

DOCTORAL THESIS

Environmental geochemistry of Potentially Toxic
Elements (PTEs) and Persistent Organic Pollutants
(POPs) as a tool of exposure evaluation and
chemical risk assessment

Matar Thiombane



Federico II University of Naples, Italy



Department of Earth Sciences, Environment and Resources (DISTAR)



Ph.D. in

Earth Sciences, Environment and Resources

XXXI Cycle

Ph.D. Thesis

**Environmental geochemistry of Potentially Toxic Elements (PTEs)
and Persistent Organic Pollutants (POPs) as a tool of exposure
evaluation and chemical risk assessment**

Candidate:

Matar Thiombane

Tutor

Prof. Stefano Albanese

Co-tutors

Prof. Benedetto De Vivo

Dr. Marcello Di Bonito

2015-2018

Table of contents

Acknowledgments.....	VII
----------------------	-----

Chapter 1: Synopsis

Introduction.....	2
Aims of the research project.....	3
Outlines of the research activities.....	4
Publications.....	16

Chapter 2: Multi-elements geochemical modelling and Potentially Toxic elements insights

Section 2.1: Geogenic versus anthropogenic behaviour and geochemical footprint of Al, Na, K and P in the Campania region (Southern Italy) soils through compositional data analysis and enrichment factor.	19
Section 2.2: Soil contamination compositional index: A new approach to quantify contamination demonstrated by assessing compositional source patterns of potentially toxic elements in the Campania Region (Italy).....	46
Section 2.3: Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) through a compositional discrimination analysis: A case study on the Campanian topsoil data.....	66
Section 2.4: Exploratory analysis of multi-element geochemical patterns in soil from the Sarno River Basin (Campania region, southern Italy) through compositional data analysis (CODA).....	86
Section 2.5: Soil geochemical follow-up in the Cilento World Heritage Park (Italy) through exploratory compositional data analysis and C-A fractal model.....	109

Chapter 3: Environmental geochemistry of organic pollutants, Human and Ecological health risk assessment

Section 3.1: Source patterns and contamination level of polycyclic aromatic hydrocarbons (PAHs) in urban and rural areas of Southern Italian soil.....	134
Section 3.2: Status, sources and contamination levels of organochlorine pesticide residues in urban and agricultural areas: a preliminary review in central-southern Italian soils.....	159
References.....	188

Appended publications.....214

Paper 1

Thiombane, M., Di Bonito, M., Albanese, A., Zuzolo, D., Lima, A., De Vivo, D. (2018). Geogenic versus anthropogenic behaviour of geochemical phosphorus footprint in the Campania region (Southern Italy) soils through compositional data analysis and enrichment factor. *Geoderma*. Volume 335, 1 February 2019, Pages 12-26. <https://doi.org/10.1016/j.geoderma.2018.08.008>.

Paper 2

Petrik, A., **Thiombane, M.,** Lima, A., Albanese, S., De Vivo. B. (2018). Soil Compositional Contamination Index: a new approach to reveal and quantify contamination patterns of potentially toxic elements in Campania Region (Italy), *Journal of applied geochemistry*. 96, 264-276. <https://doi.org/10.1016/j.apgeochem.2018.07.014>.

Paper 3

Thiombane, M., Petrik, A., Albanese, S., Lima, A., De Vivo. B. (2018). Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) through a compositional discrimination analysis: a case study on the Campanian topsoil data. *Geoderma*, 331, 87-99. <https://doi.org/10.1016/j.geoderma.2018.06.019>.

Paper 4

Thiombane M., Albanese, S., Martín-Fernández, J., Lima A., Doherty A. De Vivo, B., (2018). Exploratory analysis of multi-element geochemical patterns in soil from the Sarno River Basin (Campania region, southern Italy) through compositional data analysis (CODA). *Journal of Geochemical exploitation*, 195, 110-120. <https://doi.org/10.1016/j.gexplo.2017.06.010>.

Paper 5

Thiombane, M., Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Cavaliere, M., De Vivo.B., (2018). Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model. *Journal of Geo Exploration*. 172, 174-183. <https://doi.org/10.1016/j.gexplo.2018.03.010>.

Paper 6

Thiombane, M., Albanese, S, Di Bonito, M., Lima, A., Rolandi, R., Qi, S., De Vivo. B. (2018). Source patterns and characterisation of Polycyclic Aromatic hydrocarbons (PAHs) in urban and rural areas of Southern Italy. *Journal of environmental geochemistry and health*. <https://doi.org/10.1007/s10653-018-0147-3>.

Paper 7

Thiombane, M., Petrik, A., Di Bonito, M., Albanese, S., Zuzolo, D., Cicchella, D., Lima, A., Qu, C., Qi, S., De Vivo B., (2018). Status, sources and contamination levels of organochlorine pesticides residues in urban and agricultural areas: A preliminary review in central-southern Italian soils. *Journal of Environmental Science and Pollution Research*. <https://doi.org/10.1007/s11356-018-2688-5>.

Dedicated to my Mother

Lala Cissé (Rest in Peace)

“What you desire for yourself, wish it to others” Lala Cissé

Acknowledgments

I would like to manifest my profound gratitude to my supervisor, ***Professor Benedetto De Vivo***, for his constant guidance, supervision, advices, encouragement, support, valuable discussions and critical evaluation throughout my work; he is the brain behind the success of this project. I am also very grateful for all the practical help that my Co-supervisor, Prof. Stefano Albanese has profuse for the successful realization of my thesis.

I would like to sincerely express my appreciation to Prof. Marcello Di Bonito for his supervisor capacity during my PhD research internship at Nottingham Trent University (NTU, UK) as well as his assistance with the revision of the manuscripts. I'm equally thankful to Prof. Annamaria Lima and Prof. Domenico Cicchella for their advices and valuable discussions around the geochemical processes controlling elements patterns.

I am sincerely grateful to the Ministry of Education, Universities and Research (MIUR, Italy) for awarding to me the scholarship for the duration of this project. The financial support for the project "Industrial Research Project, BioPoliS" funded in the frame of Operative National Programme Research and Competitiveness.

I thank all the lecturers and colleagues at the Department of Earth Sciences, Environment and Resources (DISTAR), especially those in Environmental Geochemistry Laboratory. Thanks to Dr. Daniela Zuzolo, Dr. Attila Petrik, Dr. Angela Doherty, Dr. Carmela Rezza, Dr. Giulia Minolfi, Dr. Chengkai Qu, and Alberto Fortelli.

The supports of all staff at the School of Animal, Rural and Environmental Sciences, of Nottingham Trent University (NTU, United Kingdom) are greatly appreciated. I Acknowledge the Health and Wellbeing thematic research stream of the NTU for funding my field trip sampling in Senegal through the project "Bio-accessibility of mercury and Potentially Toxic Elements (PTEs) in artisanal and small-scale gold mining (ASGM), Kedougou, Senegal".

I fully wish to thank my wife Virginia Napoli for her patience, assistance and encouragements.

Finally I would like to thank my Family and everybody who was important to the successful realization of my thesis, as well as expressing my apology that I could not mention personally one by one.

Naples Italy, November 2018
Matar Thiombane

Chapter 1

Synopsis

1. Introduction

Environmental pollution is one of the most challenging environmental issues to tackle due to its impact to human health and the ecosystem. One of the main objectives of environmental geochemistry is to investigate, characterise, and reveal the patterns of organic compounds and inorganic elements and further unveil their possible sources. Geogenic features and anthropogenic activities are the main sources of environmental contamination (Lima et al., 2003a; Albanese et al., 2007; Reimann et al., 2008) which are likely to release these contaminants into atmospheric, soil and water media (Prapamontol and Stevenson, 1991; Suchan et al., 2004). Moreover, anthropogenic activities let out chemicals produced from industrial activities, domestic, livestock and municipal wastes (including wastewater), agrochemicals, and petroleum-derived products (Reimann and De Caritat, 2005; Luo et al., 2009; Bundschuh et al., 2012).

Organic pollutants cover a large group of synthesized pollutants and Persistent Organic Pollutants (POPs) have received a specific attention due to their physico-chemical properties (e.g., volatility and persistency), high toxicity, and subject to long-range atmospheric transfer (Weinberg, 1998, Szeto et al., 1991; Fang et al., 2017; Qu et al., 2018). Polychlorinated biphenyls (PCBs), Polycyclic Aromatic Hydrocarbons (PAHs) and Organochlorines Pesticides (OCPs) are the main POPs that are subject to different regulation schemes to their irreversible adverse effects to both human and wildlife health. Stockholm Convention (2005), Rotterdam and Basel (1995), World Health organisation (WHO) and United Nations Economic Commission for Europe POPs Protocol (UNECE, 2010) have so far addressed, threatened and introduced legislation which ban or fix threshold's values of these POPs into environment.

Potentially Toxic Elements (PTEs) are widespread metals/metalloids related to geogenic and/or anthropogenic activities (Lima et al., 2003a; Albanese et al., 2007; Reimann et al., 2008). PTEs are one of the major concerns in the environment because their concentrations are increasing due to accelerated population growth rate, higher level of urbanisation and industrialisation providing a great variety of anthropogenic contamination/pollution sources (Wang et al., 2012; Wu et al., 2015; Guillén et al., 2017). They have often been given special emphasis because their accumulation in different matrices can cause soil and land degradation and they can be transferred into the human body as a consequence of dermal contact, inhalation and ingestion through food chain and drinking water (Lim et al., 2008; Ji et al., 2008; Varrica et al., 2014). PTEs are generally non-biodegradable having long biological half-lives and tend to accumulate in soils being absorbed to clay minerals and organic matter (Kabata-Pendias, 2011). However, their bioavailability is influenced by different physicochemical processes (e.g. pH, Eh) and physiological adaptation (Yang et al., 2004; Skordas and Kelepertsis, 2005; Barkouch et al., 2007; Zhao et al., 2013, 2014).

PTEs and POPs can be observed in different environmental media but soil is considered an important reservoir due to its physico-chemical properties which confer high retention capacity of these pollutants. Soil contamination has been increasing worldwide (Luo et al., 2009; DEA, 2010; SSR, 2010; EEA, 2014) and has become the focus of attention in recent years. Several soil parent materials are natural sources of certain organic contaminants, elements, and these can pose a risk to the environment and human health at elevated concentrations. For that, various geostatistical computations have been used to identify source patterns of different pollutants related to underlying geological features and/or anthropogenic activities (Lima et al. 2003; Albanese et al., 2007; Reimann et al., 2008), and to further distinguish mineralisation from contamination.

Several single and complex contamination/mineralisation indices such as Enrichment Factor (Chester and Stoner, 1973), Geo-accumulation Index (Muller, 1969) or Single Pollution Index (Hakanson, 1980; Muller, 1981) have been elaborated to quantify the contamination or mineralisation status of different PTEs. They are generally based on intervention limits (thresholds) or background/baseline values of a single element based on National Legislation, as a reference. Indices based on intervention limits (thresholds) are easily interpretable and comparable, but they disregard the compositional nature of geochemical data; hence they can be biased and/or spurious (Filzmoser et al., 2009a; Pawlowsky-Glahn and Buccianti, 2011; Filzmoser et al., 2012).

This PhD research project reveals novel geostatistical computations that will lay out sources patterns of Potentially Toxic Elements (PTEs) and Persistent Organic Pollutants (POPs), and assess the soils contamination levels in the central-southern Italy. Series of follow up studies have provided an invaluable baseline for these contaminants distribution in Italy to push towards an institutional response for more adequate regulation of these pollutants worldwide.

2. Aims of this PhD research programme

The main aim of this research study is to investigate sources patterns and contamination levels of POPs and PTEs in soil of urban and rural areas of various regions. During my PhD research programme, seven peer-reviews have been published, where we investigated:

- I. Possible sources and contamination levels of 24 OCPs compounds in urban and rural soils from central and southern Italy.
- II. The characterisation, possible sources and toxicity equivalent quantity level of sixteen PAHs in urban and rural soils of southern of Italy.
- III. The development of a new comprehensive discrimination analysis to identify possible sources of contamination or/and enrichment of four PTEs (Zn, Pb, Cr and Ni) in the Campania region.

- IV. The compositional behaviour of 15 PTEs in Campania soils and computed a new Soil Compositional Contamination Index (SCCI) that highlights contamination levels.
- V. The potential impacts and footprint of some PTEs related to the agricultural activities in Campania region.
- VI. The potential geochemical patterns and contamination sources of PTEs in soils of the Sarno River Basin (Campania, Southern Italy).
- VII. The distinction between geogenic to anthropogenic geochemical signatures associated to anomalies into the World Heritage Territory known as “National Park of the Cilento and Vallo di Diano” (Southern Italy).

A further ongoing research project is currently investigating the content and bioavailability of mercury and Potentially Toxic Elements (PTEs) in artisanal and small-scale gold mining (ASGM) districts of Kedougou (Senegal). This study in particular will represent a fundamental stepping stone to build a baseline review of PTEs in ASGM of Kedougou (Senegal) and evaluate human health risks from exposure of PTEs. It is envisaged that the results of this study should trigger more detailed surveys in contaminated areas as well as ad-hoc risk-based studies, which in the long-term will constitute a strong argument to cause an adequate institutional response by the Senegalese regulating authorities for a full application the Minamata convention.

3. Outline of the thesis

The summary of this thesis is based on studies reported in 7 appended peer-reviewed papers and an ongoing research project.

3.1. Sources patterns and contamination levels of POPs and PTEs in soils of the central-southern Italy

The total survey area (considering administrative regional boundaries) extended to 157,716 km² with 31.26 million of inhabitants, mostly grouped in main urban areas (ISTAT, 2016). The survey area included 4 administrative regions (Latium, Marche, Tuscany and Umbria) from central and 7 (Abruzzo, Apulia, Basilicata, Calabria, Campania Molise, and Sicily) from southern Italy (Fig. 1).

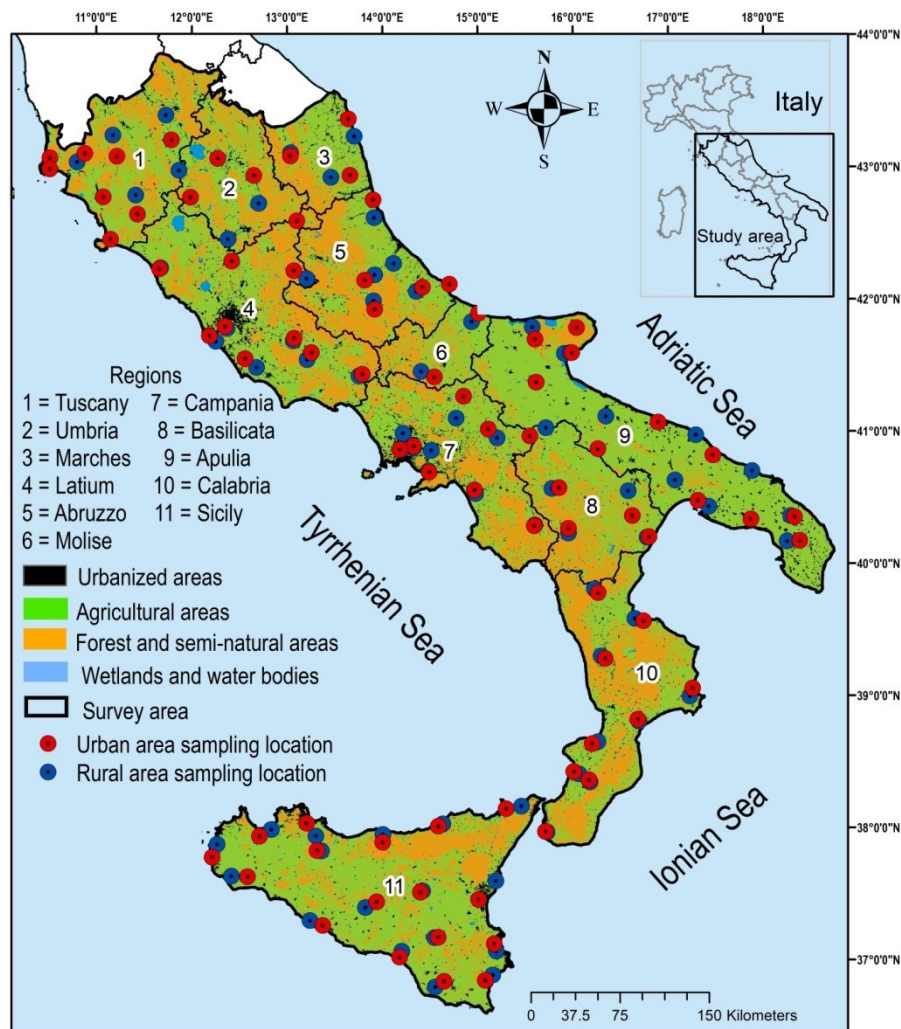


Fig. 1. Land use/land cover of the study area (simplified from Corine Land Cover 2012). Urban (red dots) and rural (blue dots) sampling site locations are displayed.

Most of the land is used for agricultural and forestry activities. Agriculture occupies one-fourth of the land available, which includes cultivation of hilly areas where agriculture results in modifying the natural landscape and resources through terracing, irrigation, and soil management (Corona et al., 2012; ISTAT, 2013; ISPRA, 2014a). Large urban areas such as Rome (Latium), Naples (Campania), Bari (Apulia) and Palermo (Sicily), are densely populated and surrounded by metropolitan areas where both industrial activities, manufactories and intensive agriculture occur (ISTAT, 2016).

Based on the land cover, industrial and main urbanised areas, a specific sampling site was designed. The sampling campaign took place from early April to end of September 2016, with the aim to select the most representative topsoil samples in urban and rural areas throughout 11 regions (Latium, Marche, Tuscany, Umbria, Abruzzo, Apulia, Basilicata, Calabria, Campania, Molise, and Sicily) from the centre to southern Italy. In each region, the main urban areas and the nearest rural areas where most of the land is devoted to agricultural activities, were selected. Site selection was carried out by interpreting, using Geographical Information Systems (ArcGIS™, 2012), information on land use/land cover of the study

area (ISPRA, 2014b; Corine land cover, 2012) together with satellite imagery (Google Earth® professional, 2016). A total of 148 soil samples were collected with a nominal density of 2 samples/ 2500 km² (in urban and rural areas) (Fig. 1).

Samples have been collected from public gardens in urban areas, and from agricultural land (farmlands/cropland) in rural areas (Fig. 2). All the samples were collected using a stainless steel scoop, kept in labelled glass bottles and directly stored in ice boxes to minimize the losses caused by volatilization and initial degradation of the organic compounds. Each topsoil sample (from 0-20 cm) was made by homogenizing 5 subsamples at the corners and the centre of a 100m² square, collecting approximately 1.5 kg in total. The sampling procedure followed the Geochemical Mapping of Agricultural and Grazing Land Soil (GEMAS) sampling procedure described by Reimann et al. (2014).



Fig. 2. Characterisation of the sampling sites. Captions (A) and (B) shows sampling location in agricultural land and sampling protocol, respectively; (C) corresponds to a urban sampling site whereas (D) reveals sampling location depth (20cm).

Soil samples were homogenized and sieved using a <2mm mesh sieve after removing stones, detritus and residual roots. Finally, composite samples were stored at -4 °C in the environmental geochemistry laboratory of the University of Naples Federico II (Italy) until instrumental analysis of organic compounds. Geographical coordinates were recorded by geospatial positioning systems (GPS, using the WGS84 model) at each sample site.

Inorganic elements were analysed at an international accredited Laboratory, Acme Analytical Laboratories Ltd (now Bureau Veritas, Vancouver, Canada). The samples were analysed after the ultra-trace aqua regia extraction method, by inductively coupled plasma mass spectrometry (ICP/MS) for “pseudototal” concentration of 53 elements. The “pseudototal” corresponds to the part concentration of elements extractable using strong acid solution (aqua regia digestion). It estimates the maximum amounts of metals that could hypothetically be mobilized and transported in the environment from geogenic part and all anthropogenic inputs (Vercoetere et al., 1995; Adamo and Zampella, 2008). Although primary silicates are not dissolved, metals associated with most other major soil components are liberated. The 2-mm sieved soil was characterised according to the Italian official methods of soil analysis (Violante and Adamo, 2000). A sub-sample of 15 g was digested in 90 ml aqua regia and leached for 1 hour in a 95°C water bath. After cooling, the solution was diluted to a final volume of 300 ml using a solution of 5% HCl. The sample weight to solution volume ratio was 1 g per 20 ml. The solutions were analysed using a Perkin Elmer Elan 6,000/9,000 inductively coupled plasma emission mass spectrometer (ICP-MS). The accuracy and precision of the data was measured by comparison to known analytical standards. Calibration solutions were included at the beginning and the end of each analytical run (a total of 40 solutions). Precision is $\pm 100\%$ at the detection limit, and improves to better than $\pm 10\%$ at concentrations 50 times the detection limit or greater.

Targeted organic compounds were organochlorine pesticides (OCP), Polycyclic Aromatic Hydrocarbons (PAHs) and Polychlorinated Biphenyl (PCB), and soil samples were analysed at the Key Laboratory of Biogeology and Environmental Geology of Ministry of Education at the University of Geosciences in Wuhan, China. PAHs and PCB were analysed using GC-MS (Agilent 6890N/5975 MSD) coupled with a HP-5972 mass selective detector operated in the electron impact mode (70 eV) installed with a DB-5 capillary column (30m×0.25 mm diameter, 0.25µm film thickness). Helium (99.999%) was used as the GC carrier gas at a constant flow of 1.5 mL/ min. An 1µL concentrated sample was injected with splitless mode. The chromatographic conditions were as follows: injector temperature 270 °C; detector temperature 280 °C; oven temperature initially at 60°C for 5 min, increased to 290 °C at 3 °C/ min, and held for 40 min. Chromatographic peaks of samples were identified by mass spectra and retention time. Twenty congeners OCPs concentrations were carried out by an Agilent 7890A gas chromatograph

with a ^{63}Ni electron capture detector (GC-ECD) equipped with a DB-5 capillary column (30.0 m length, 0.32 mm diameter, 0.25 mm film thickness). Gas chromatography-mass spectrometry (GC-MS) and gas chromatography-electron capture detector (GC-ECD) are the most common and appropriate systems to investigate organic contaminants in different environmental media. Many authors (e.g., Aramendia et al., 2007; Alves et al., 2012) showed the high sensitivity of GC-ECD for organophosphorus and organochlorine pesticides. In this study, the rationale of working with GC-ECD analyser was based on the excellent sensitivity and satisfactory quantification limits, allowing the identification and quantification of pesticides at low levels. A 10 g of dried soil samples were spiked with 20 ng of 2,4,5,6-tetrachloro-m-xylene (TCmX) and decachlorobiphenyl (PCB209) as recovery surrogates and were Soxhlet-extracted with dichloromethane for 24 h. Activated copper granules were added to the collection flask to remove elemental sulphur. The extraction of OCPs was concentrated and solvent-exchanged to n-hexane and further reduced to 2–3 mL by rotary evaporation. The alumina/silica (1:2) gel column (450°C muffle drying for 4 h, both deactivated with three percent water) was used to purify the extract and OCPs were eluted with 30 mL of dichloromethane/hexane (2/3). Then the eluate was concentrated to 0.2 mL under a gentle nitrogen stream and a known quantity of penta-chloronitrobenzene (PCNB) was added as an internal standard prior to gas chromatography–electron (GC–ECD) analysis. Nitrogen was used as carrier gas at 2.5 mL/min under constant-flow mode. Injector and detector temperatures were maintained at 290°C and 300°C, respectively. The oven temperature started from 100°C (with an equilibration time of 1 min), and rose to 200°C at a rate of 4°C/min, then to 230°C at 2°C/min, and finally reached 280°C at 8°C/min, and was held for 15 min. 2 μL of each sample was injected into the GC- μECD system for the analysis. Concentration of the individual target OCPs were identified by comparison of their retention times (previously confirmed with GC/MS) and quantified using an internal standard. The gas chromatograph (GC-MS) parameters of the Agilent 6890GC-5975MSD system were the same as those of the Agilent 6890 GC equipped with ^{63}Ni micro-electron capture detector (GC- μECD). The mass spectrometer (MS) was operated in electron impact ionization mode with electron energy of 70 eV. The ion source, quadrupole and transfer line temperatures were held at 230, 150 and 280°C, respectively. Target compounds were monitored in selected ion monitoring (SIM) mode.

Procedure types used for quality assurance and quality/control (QA/QC) were as follows: method blank control (procedural blank samples), parallel sample control (duplicate samples), solvent blank control, and basic matter control (US EPA, 2000). The spiked samples containing internal standard compounds were analysed simultaneously with soil samples. A procedural blank and a replicate sample were run with every set of 12 samples analyzed to check for contamination from solvents and glassware. The limits of detection (LODs) were based on 3:1 S/N ratio. TCmX and PCB 209 were spiked as surrogate standards to judge procedural performance. The surrogate recoveries for TCmX and PCB209 were 77.8

$\pm 19.0 \%$ and $89.3 \pm 20.3 \%$, respectively. The relative standard deviation (RSD) was less than 10%. All OCPs concentrations were expressed on an air-dried weight basis (ng/g).

This study is important because it will represent a fundamental assessment that shows and help to build a long-overdue national picture of PTEs and POPs status in Italian soils. More detailed surveys will be followed up in contaminated areas, which would constitute a baseline to promote an adequate institutional response by the Italian regulating authorities.

3.2. Bio-accessibility of mercury and Potentially Toxic Elements (PTEs) in artisanal and small-scale gold mining (ASGM), Kedougou, Senegal

Environmental geochemistry aims to reveal the sources and contamination level of inorganic elements in different environmental media, released from geogenic features and anthropogenic activities. Potentially Toxic Elements (PTEs) occur in soils, sediments, air and can bio-accumulate in food chains, reaching human body (Gosar et al., 2003; Campbell et al., 2005; Ottesen et al., 2013; Niane et al., 2014). Some human activities may perturb or alter natural cycles of metals in the environment leading to accumulation of many PTEs in the food chain. Among PTEs, Hg has attracted the public attention in this last decade, because of its high toxicity level, harmfulness to human health and wildlife (Veiga et al., 2004; WHO, 2013). Over the last 30 years, Hg has been much investigated regarding its impact in the ecosystem due to its mobility and volatility, and its potential methylation and bioaccumulation. No essential biological function of Hg is known. On the contrary, Hg is among the most toxic elements to human and wildlife. All chemical forms of Hg are toxic to humans (Winship, 1986; Roi and Sabbiani, 1993; Clarkson, 1997). Inorganic forms of Hg show acute toxicity, with a variety of symptoms and damage to organs. Some organomercurial, in particular low molecular-weight alkyl compounds, are considered even more toxic to humans because of their high chronic toxicity with respect to various, largely irreversible, defects of the nervous system (Clarkson and Magos, 2006). Methylmercury (CH_3Hg) is particularly significant in this respect because it is produced in nature by microorganisms where reducing conditions prevail. Methyl Hg shows strong teratogenic effects, and carcinogenic and mutagenic activity has also been implied.

The most significant anthropogenic activities giving rise to past and present emissions of Hg to land, water, and air appear to be the following:

- (1) Mining and smelting of ores, in particular Cu and Zn.
- (2) Combustion of fossil fuels, mainly coal.
- (3) Industrial production processes, in particular the Hg cell chlor-alkali process and cement production.
- (4) Consumption-related discharges, including waste incineration.

(5) Gold production using Hg-amalgamation technology in artisanal and small-scale Au mining (ASGM). Artisanal, or small scale Au mining (ASGM) is one of the most significant anthropogenic sources of Hg and other potentially toxic elements (PTE) release into the environment, in developing countries and countries with economies in transition, as at least a quarter of the world's total Au supply comes from such sources (Pirrone et al., 2010). Telmer and Veiga (2008) revealed that ASGM activities release between 640 to 1350 Megagrams (Mg) of Hg per annum into the environment, averaging 1000 Mg Hg/yr, from at least 70 countries. 350 Mg Hg/yr of this are directly emitted to the atmosphere while the remainder (650 Mg/yr) is released into the hydrosphere (rivers, lakes, soils, tailings). As pointed out by the World Health Organisation (WHO) guidance for identifying populations at risk from Hg exposure (WHO, 2003) and Minamata convention (UNEP 2013), some of these activities are particularly common in developing countries, where regulations, environmental controls and monitoring tend to be limited in amplitude and resources, producing 'hot spots' of risk from exposure. Mining activities can release large amount of PTEs in the environment media such as groundwater, soils and air. Furthermore, Hg has bio-accumulation propriety and can reach the food chain (vegetables and animals). This contaminant persists in nature in different chemical forms like methylmercury (methylHg) which constitute a real concern for human and wildlife health. The aims of this study are to:

- (1) Assess the distribution and baseline of PTEs in soils and sediments of the Kedougou Gold mining district (Senegal);
- (2) Quantify the total, bioavailable, and bio-accessible portions of Hg and its various chemical species in different environmental media and in food stuff;
- (3) Identify the main factors influencing the fate and mobility of contaminants through different media;
- (4) Assess human health risk amongst affected population groups.

This study will be the first step forward study on looking at human health and a toxicological assessment of Hg and PTEs in ASGM in Kedougou. Different environmental media have been taken into consideration and geostatistical computations will help to assess the spatial distribution and sources patterns of PTEs in our study area.

The Kedougou Region is situated in the south-eastern part of Senegal and occupies a territory of about 16,800 km². The region borders Mali and the Guinee Bissau at the eastern and southern sides, respectively. The main geological features of the region are constituted essentially of Precambrian basement formed by the so-called Kedougou-Kénieba inlier (Fig. 3) (Bassot, 1997). This consists of volcanic, volcano-sedimentary and sedimentary formations, which are distributed into two Supergroups: the Mako Supergroups or Mako Volcanic Belt to the West and the Dialé-Daléma Supergroups in the East (Fig. 3). The geological formations from these two Supergroups serve as hosts to several

generations of granitoids in coalescent solid masses distributed into two batholiths, Badon-Kakadian and Saraya, intruding the Mako and Dialé-Daléma Supergroups, respectively.

The Falémé Volcanic Belt crops out to the east of the Daléma basin, dominated by plutonic rocks, consisting of two plutonic complexes, the Balangouma pluton and the Boboti pluton in the centre and south of the belt, respectively.

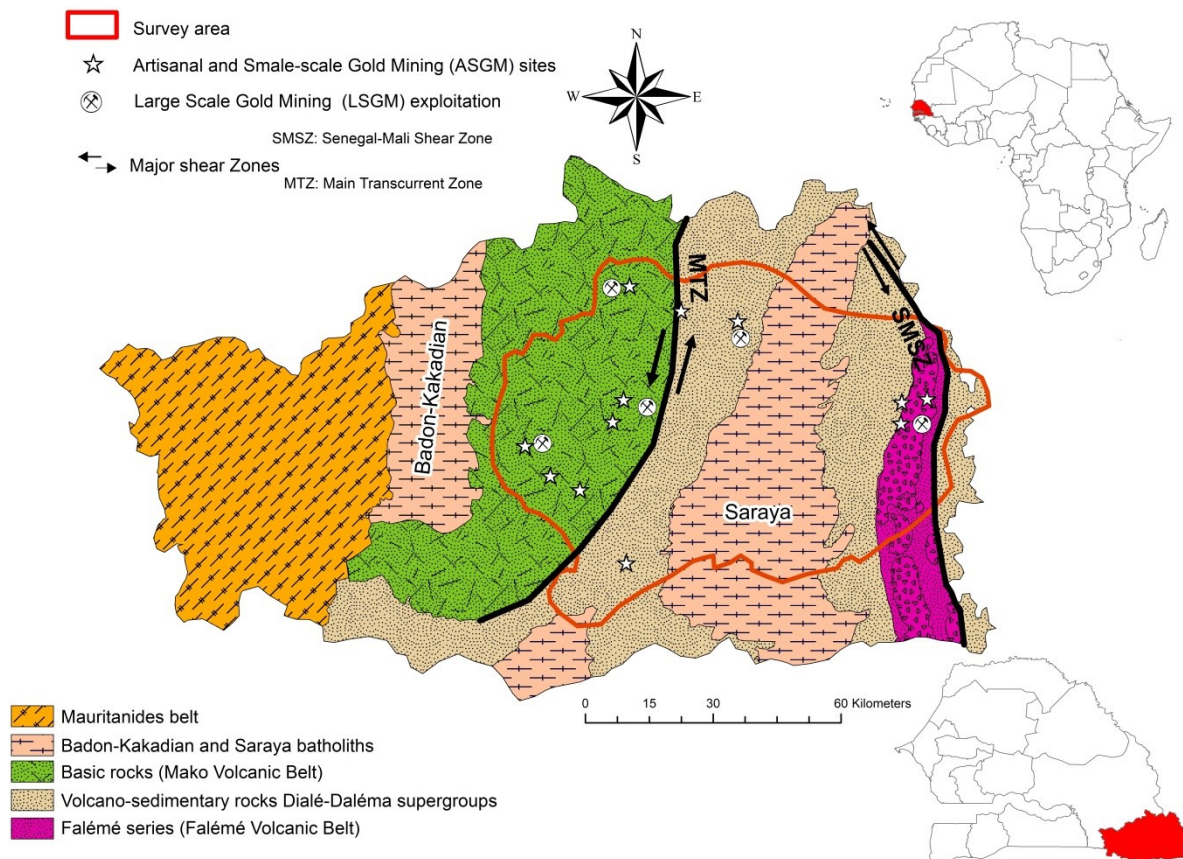


Fig. 3. Geological features of the survey area (modified from Bassot 1997). In the map are displayed location of Large (LSGM) and artisanal small (ASGM) gold mining districts

These ancient lands, commonly called "Birimian formations" constitute a metallogenic province of great ore deposits such important Au-bearing reserves. The main Au deposits in the Kedougou region are related to the major tectonic zones namely: the Senegalo–Malian Shear Zone (SMSZ) and the Main Transcurrent Zone (MTZ) (Fig. 3), which are major regional controls of Au mineralisation (Bassot, 1997; Sylla and Ngom, 1997). There are two types of Au occurrences: alluvial Au and Au-bearing quartz veins hosted by shear zones. The Kedougou region is an emerging significant Au camp where Teranga Gold Operation (TGO) is currently the largest active mining company exploiting shear zone hosting Au-bearing quartz veins.

Several Au mining companies are in prospection steps such Randgold-Sénégal company, Toro Gold, and Oromine company. In addition to this, a large number of ASGM activities are scattered throughout the Kedougou region (Fig. 4).

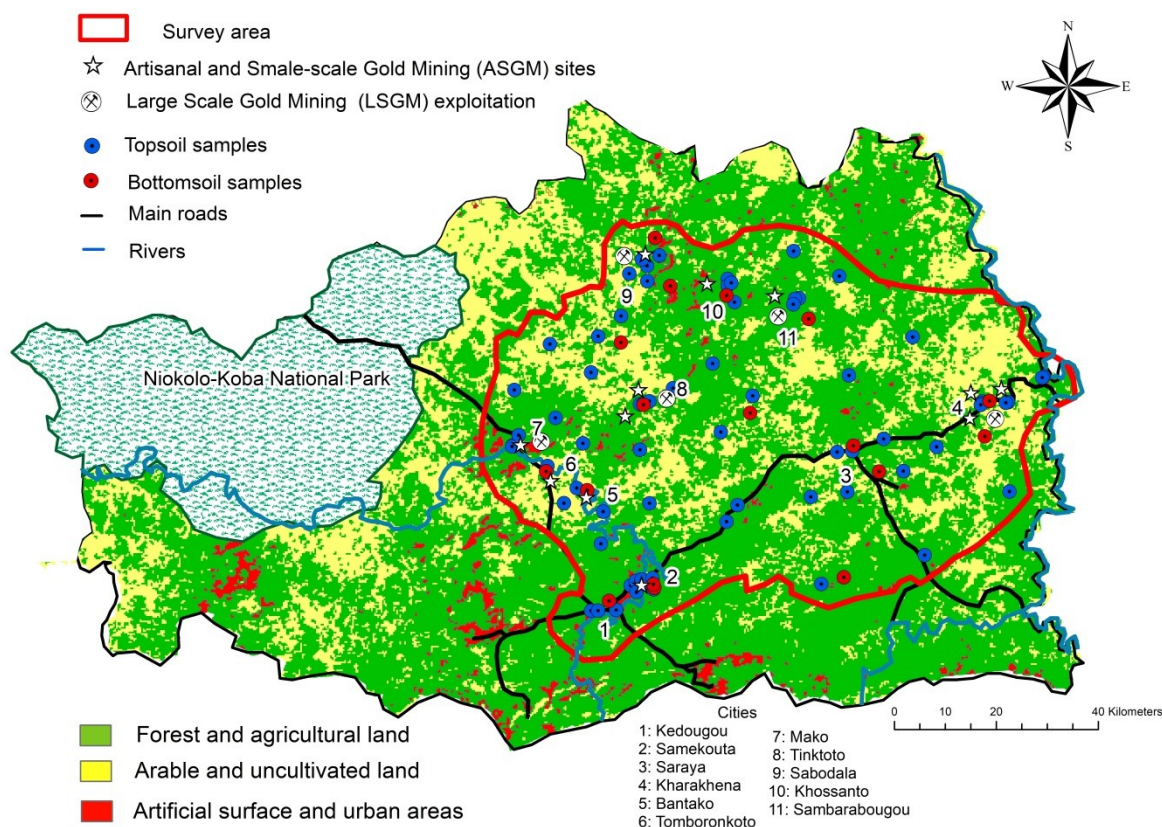


Fig. 4. Land-cover/land-use and sample locations in the study area.

The relief of the region is the most rugged of the country with large number culminating points more in the Southern part. It should be noted that tropical ferruginous soils and poorly evolved soil erosion predominate in the region. Such a situation gives it a very important biodiversity both from a faunal and a floristic point of view. Indeed, one meets almost all the species of the Guinean flora present in Senegal but also a good part of the fauna of West Africa. This biodiversity is partly the result of favourable climatic characteristics. The region is one of the last strongholds of Senegal's fauna composed of the most prestigious antelope bands of the world as well as buffaloes, lions, panthers, crocodiles mostly in the biggest natural park of the country, the Niokolo-Koba National Park (Fig. 4).

The hydrographic network is dense and depends very heavily on rainfall. The area is irrigated by the Falémé and the Gambia River and its large number of tributaries. The Gambia and Falémé, which constitute the two major rivers, are at low water levels and play an important role in local development (agriculture, water supply for men and livestock) (Fig. 4).

With a population of 156,352 inhabitants in 2017, the region remains sparsely populated, representing only 1.2% of the population of Senegal with a density of 9.1 inhabitants per km².

The Kedougou region has significant mining potential, making it a magnet for investors. The geological substratum of the region consists essentially of Precambrian basement. Mining activities have boosted the craft sector which was very embryonic during the latter years. Gold mineralisation is the main ore deposit that is being exploited industrially and through ASGM.

The National Agency of statistics and demography of Senegal (ANDS 2015) has outlined 29 ASGM activities settling inner Kedougou and most of them practice Au-Hg-panning activities in the bank of Gambia and Falémé rivers (Fig. 4). Moreover, river-Hg-contamination may constitute a real concern for population living in these areas consuming fishes from the rivers (Counter and Buchanan, 2004; Ishikawa and Ikegaki, 1980). In fact, Hg from geogenic and anthropogenic (e.g. ASGM for Au amalgamation) activities may accumulate in sediments through runoff and inflow, wet and dry deposition. In anaerobic media, bacteria can transform Hg through a methylation and CH₃Hg could bioaccumulate in the biota and go through the food chain. This organomercurial is very toxic to human and is considered neurotoxic and teratogenic (WHO 2003; UNEP 2013).

Based on the land-use and the location of large and small Au mining sites, a specific sampling design was planned. The fieldwork was organised in overall March 2018 aiming to collect most representative soils, sediments, green vegetables, fishes and human's hair samples in the study area (Fig. 5).

A Total of 83 soil samples from topsoil and 20 bottom soils (1-5 m) were collected in 6,742 km² of the study area, following a specific sampling density. A thick soil sample density was planned in surrounding LSGM and ASGM districts, and a lower one in correspondence of unused or arable land (Fig. 4). Each top soil sample (from 0-20 cm) was made by homogenizing 5 subsamples at the corners and the centre of a 100m² square, collecting approximately 1.25 kg in total. This sampling procedure follows the Geochemical Mapping of Agricultural and Grazing Land Soils (GEMAS) sampling procedure (Reimann et al., 2008). Bottom soil samples are collected in depth from 1-5 m and mostly in ASGM sites. These subsoil samples will allow assessing the background and baseline concentrations of PTEs in the survey area.

Seventeen (17) sediment samples were collected at the bank of the Gambia and Falémé rivers, particularly in parts of the rivers nearby ASGM sites (Fig. 5B). At each sampling site, a top-layer (10-30 cm) of the sediment was collected, measuring as well physico-chemical properties (e.g. including pH, conductivity, total dissolved oxygen) and the geographical coordinates system recorded by geospatial positioning systems (GPS).



Fig. 5. Sampling procedures of different sample matrices. (A) Characterisation of soil samples location; (B) sediment sampling procedure in the bank of the Gambia River; (C) measurement of length and weight of the fish sample (*Hepsetus odoe*); (D) Vegetable garden sampling site along the Gambia River.

Biological samples are considered useful markers to investigate contamination levels and can constitute direct tools to assess human health and ecological risks. In our study we were focused on vegetables, fishes and human's hair samples to retrace speciation of PTEs in different environmental media and the pathways that induce their transfers.

Length and weight of the fish samples were measured as well as the characterisation of their corresponding immediate environment of the different species (Fig. 5C). Furthermore, muscle (flesh) samples were collected above the lateral line, between the dorsal fin and the caudal fin. This will maximise the amount of muscle tissue collected and reduce the risk of accidentally piercing internal organs. Cutting below the fish's lateral line ensure the lower intestine or other internal organs are not pierced. If the intestine is cut open, this will lead to contamination of the organs and the sample will not be usable. Muscle samples were placed in labelled storage container/bag and stored at -4°C before analytic procedures. This biological sample preparation followed the manual of Monitoring and Sampling for biological assessment "Fish collection and dissection for the purpose of chemical analysis of tissues" of the Department of Environment and Heritage protection of UK (2017).

Vegetal samples were collected in gardens located along the rivers (Fig. 5D). In this specific sampling design, we aimed to assess the transfer of Hg and potentially toxic elements into vegetables which are directly irrigated from the rivers. In addition, these vegetable gardens are located in villages near ASGM sites. Two vegetable species from the Asteraceae (Lettuce: *Lactuca sativa* and Cabbage: *Brassica oleracea*) were selected and 10 samples of each were collected. Each sample was rinsed and thoroughly washed with tap water to remove dust particles, placed in plastic bag and frozen as soon as possible. After collection, the vegetable samples were taken to the laboratory for preparation. Specifically, the edible part of each lettuce was removed with a ceramic knife, weighed and then oven-dried at 60° C.

Scalp hairs were collected to voluntary peoples at hospital and dispensary, across the whole study area, from people willing to make their hair available after receiving detailed information on the aims of the current study. This project targeted three receptor groups such adult miners (they must be involved at least 5 years in artisanal Au mining and leaving in the survey area), children (between 0-5 years) and women. Specifically, a 10 g hair sample was collected from the scalp of each person, by means of stainless steel scissors, and stored in a paper bag, properly sealed. For each collected sample gender, age and municipality of residence of the donors were recorded. All the collected samples were taken to the sample preparation laboratory at the British Geological Survey (BGS, United Kingdom) for chemical analysis.

Chemical analyses of soil and sediments were carried out at an international accredited Laboratory, Acme Analytical Laboratories Ltd (now Bureau Veritas, Vancouver, Canada). The samples were analysed after an aqua regia extraction, by inductively coupled plasma mass spectrometry (ICP/MS) for “pseudototal” concentration of 53 elements. The measurement of inorganic contaminant bio-accessibility from soil and sediment is based in agreement with the in vivo validated Unified BARGE Method (Caboche, 2009). Biological samples (Vegetable, Fish and Human Hair) are analysed at the environmental geochemistry laboratory of the British Geological survey (BGS, United Kingdom) (in progress).

LIST OF PUBLICATIONS

- I. **Thiombane, M.**, Di Bonito, M., Albanese, A., Zuzolo, D., Lima, A., De Vivo, D. (2018) Geogenic versus anthropogenic behaviour of geochemical phosphorus footprint in the Campania region (Southern Italy) soils through compositional data analysis and enrichment factor. *Geoderma*. Volume 335, 1 February 2019, Pages 12-26. <https://doi.org/10.1016/j.geoderma.2018.08.008>.
- II. Petrik, A., **Thiombane, M.**, Lima, A., Albanese, S., De Vivo, B. (2018) Soil Compositional Contamination Index: a new approach to reveal and quantify contamination patterns of potentially toxic elements in Campania Region (Italy), *Journal of applied geochemistry*. 96, 264-276. <https://doi.org/10.1016/j.apgeochem.2018.07.014>.
- III. **Thiombane, M.**, Petrik, A., Albanese, S., Lima, A., De Vivo, B. (2018) Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) through a compositional discrimination analysis: a case study on the Campanian topsoil data. *Geoderma*, 331 (1), 87–99. <https://doi.org/10.1016/j.geoderma.2018.06.019>.
- IV. **Thiombane M.**, De Vivo B., Albanese S., Martín-Fernández J., Lima A., Doherty A. (2018) Assessment of the behaviour of potentially toxic elements (PTEs) in soil from the Sarno River Basin through a compositional data analysis. *Journal of Geochemical exploration*. 195, 110-120. <https://doi.org/10.1016/j.gexplo.2017.06.010>.
- V. **Thiombane, M.**, Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Cavaliere, M., De Vivo, B., (2018). Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model. *Journal of Geo Exploration*. 172, 174–183. <https://doi.org/10.1016/j.gexplo.2018.03.010>.
- VI. **Thiombane, M.**, Albanese, S., Di Bonito, M., Lima, A., Rolandi, R., Qi, S., De Vivo, B. (2018) Source patterns and characterisation of Polycyclic Aromatic hydrocarbons (PAHs) in urban and rural areas of Southern Italy. *Journal of environmental geochemistry and health*. <https://doi.org/10.1007/s10653-018-0147-3>.
- VII. **Thiombane, M.**, Petrik, A., Di Bonito, M., Albanese, S., Zuzolo, D., Cicchella, D., Lima, A., Qu, C., Qi, S., De Vivo B., (2018) Status, sources and contamination levels of organochlorine pesticides residues in urban and agricultural areas: A preliminary review in central-southern Italian soils. (*Journal of Environmental Science and Pollution Research*. Volume 25, Issue 26, pp 26361–26382. <https://doi.org/10.1007/s11356-018-2688-5>.

Chapter 2

Multi-elements geochemical modelling and Potentially Toxic elements insights

Section 2.1

Geogenic versus anthropogenic behaviour and geochemical footprint of Al, Na, K and P in the Campania region (Italy) soils through compositional data analysis and enrichment factor

Section 2.2

Soil contamination compositional index: A new approach to quantify contamination demonstrated by assessing compositional source patterns of potentially toxic elements in the Campania Region (Italy)

Section 2.3

Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) through a compositional discrimination analysis: A case study on the Campanian topsoil data

Section 2.4

Exploratory analysis of multi-element geochemical patterns in soil from the Sarno River Basin (Campania region, southern Italy) through compositional data analysis (CODA)

Section 2.5

Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model

Section 2.1

Geogenic versus anthropogenic behaviour and geochemical footprint of Al, Na, K and P in the Campania region (Southern Italy) soils through compositional data analysis and enrichment factor

This section has been published in:
Journal of Geoderma
Volume 335, 1 February 2019, Pages 12-26

Geogenic versus anthropogenic behaviour and geochemical footprint of Al, Na, K and P in the Campania region (Southern Italy) soils through compositional data analysis and enrichment factor

Abstract

Geochemical studies that focus on environmental applications tend to approach the chemical elements as individual entities and may therefore offer only partial and sometimes biased interpretations of their distributions and behaviour. A potential alternative approach is to consider a compositional data analysis, where every element is part of a whole. In this study, an integrated methodology, which included compositional data analysis, multifractal data transformations and interpolation, as well as enrichment factor analysis, was applied to a geochemical dataset for the Campania region, in the south of Italy, focusing in particular on the behaviour, footprints and sources of a smaller pool of elements: Al, Na, K and P. The initial dataset included 3669 topsoil samples, collected at an average sampling density of 1 site per 2.3 km², and analyzed (after an aqua regia extraction) by a combination of ICP-AES and ICP-MS for 53 elements. Frequency based methods (CIR biplot, Enrichment Factor computation) and frequency spatial-method (fractal and multifractal plots) allowed identifying the relationships between the elements and their possible source patterns in Campania soils in relation to a natural occurring concentrations in geogenic material (rocks, soils and sediments) or human input. Results showed how the interpretation of concentration and behaviour of Al, Na, K and P was enhanced thanks to the application of data log-ratio transformation in univariate and multivariate analysis compared to the use of raw or log-normal data. Multivariate analyses with compositional biplot allowed the identification of four element associations and their potential association with the underlying geology and/or human activities. When focusing on the smaller pool of elements (Al, P, K and Na), these relationships with the unique geology of the region, were largely confirmed by multifractal interpolated maps. However, when the local background was used for the calculation of the enrichment factor, the resulting interpolated maps allowed to identify smaller areas where the greater concentrations of P could not be possibly associated to a mineralisation (e.g., ultrapotassic rocks) but were more likely to be associated to anthropogenic input such as agriculture activities with potentially extensive use of phosphate fertilizers. The integrated approach of this study allowed a more robust qualitative and quantitative evaluation of elemental concentration, providing in particular new and vital information on the distribution and patterns of P in soils of the Campania region, but also a viable, more robust, methodological approach to regional environmental geochemistry studies.

1. Introduction

Element distributions in soils are generally related to a variety of factors such as geology, chemical reactivity, mineralogy, hydrology, vegetation, and anthropogenic activities. In recent years environmental geochemistry has allowed to gain much insight on the relationship between chemical elements in soils and their sources by means of various tools. In particular, univariate and multivariate analysis (Otero et al., 2005; Reimann et al., 2008; Zuo, 2011; Thiombane et al., 2018a) as well as spatial analysis (frequency space-method) have helped to discriminate between anthropogenic contribution (and contamination) compared to natural or geogenic sources. More recently, different approaches have also been used to treat compositional geochemical data in a more comprehensive way. Geochemical data are closed number systems (parts of a whole) and they should be treated as that to avoid spurious correlations and misleading interpretations. By taking only the absolute values or the log-normal data, the proportional nature of the geometry of the simplex may be not fully captured (Han and Kamber, 2001). Compositional data instead, consider each element as part of a whole which carry relative information (Aitchison, 1986; Egozcue et al., 2003). For this reason, and to minimize and/or eliminate the presence of outliers and spurious correlation (Pawlowsky-Glahn and Buccianti, 2011), they have increasingly been approached in alternative ways. In particular, the use of log-ratio transformations such as additive log-ratio (alr), centered log-ratio (clr) (Aitchison, 1986) and isometric log-ratio (ilr) (Egozcue et al., 2003) were found to be effective approaches to deal with these complex datasets and better explain their significance. Furthermore, mapping remains a useful tool to display data distribution through fractal and multifractal analysis (Mandelbrot, 1983; Cheng et al., 1994) and through geostatistics (Olea et al., 2018; Thiombane et al., 2018a). These have been demonstrated to be a powerful means for identifying geochemical anomalies (Cheng, 1999, 2007; Cheng et al., 2000, 2010; Lima et al., 2003a; Cicchella et al., 2005) and enhance elements patterns. Aside from their inherent complexity, often environmental geochemistry studies still present some discrepancies in the definition of key concepts and therefore interpretation of results. For example, the concept of background and/or baseline concentration are still confused and misused, whilst they should be clearly defined and interpreted: a background value corresponds to the range of concentration of a given element in a given area which is completely dependent on the compositional and mineralogical characteristic of the parent/source geological material (Reimann et al., 2005; Albanese et al., 2007). On the other hand, a baseline value relates the actual most diffused range of concentration of a given element in a specific area depending both on the nature of the parent geological/source material (Salminen and Gregorauskiene, 2000) and on the historic diffuse release into the environment from anthropogenic sources.

In reality, in areas where the anthropogenic impact is very small or not significant, the background and baseline values can overlap, as it is often difficult to distinguish the natural sources compared to the anthropogenic contribution. In areas where the anthropogenic impact is evident or significant, however, the background and baseline values should be interpreted appropriately and in a distinct way. This study focuses on a region in south of Italy, Campania, characterized by complex geological and geomorphological features and by a significant presence of human activities, such as industry, agriculture, tourism and urbanization. The industrial presence is often related to the agricultural activities and is mostly developed in proximity of the agricultural and surrounding urban areas. Many of these industrial activities are devoted to food production and preservation process (e.g., San Marzano tomatoes conserves), but also to clothes productions and tanneries (e.g., Solofra tanneries industries). These activities represent the main livelihood for the local communities and a major economic input not only for the Campania region, but for the Italian economy. Nevertheless, as with any agriculture activity, it is inevitable that chemical fertilizers (phosphates and sulfates based) are used to improve soil productivity and quality. Some of the benefits of using fertilizers, however, carry also 'externalities' that can travel from the soil, to the crop, and the entire food chain. In particular, metallic impurities of P-based fertilizers containing Cu, As and Cd, can affect the quality of the soil and possibly its contamination; phosphate fertilizers are in fact known as a major source of trace metals among all mineral fertilizers (Nziguheba and Smolders, 2008). These sources of contamination can be investigated by means of Enrichment Factor (Chester and Stoner, 1973) which is able to display the degree of contamination related to a mineralisation (geogenic) or anthropogenic activities (Reimann and de Caritat, 2005).

This study proposes an assessment of the potential impacts and footprint of some of the agricultural activities in Campania, by use of compositional data analysis, concentration-area (C-A) and spectrum areas (S-A) fractal and multifractals models applied to geochemical data. In particular, the study focuses on a smaller pool of elements that can be potentially directly related to either geogenic, anthropogenic or mixed source: Al, Na, K, and P will be investigated to detect their potential origin and their background concentrations in Campania soils. The main objectives of this study were:

(1) to illustrate the importance of compositional log-transformations on geochemical data; (2) to delineate the main sources of elements by using a combination of multivariate analysis and GIS based approach; (3) to use enrichment factor (EF) to assess geogenic or anthropogenic behaviour of P in the study area; (4) to prove the importance and ease of using the local reference elements as opposed to the traditional continental crust references values in EF calculation to determine degree of contamination. The results from this investigation could greatly influence and provide a blueprint to future similar studies.

2. Materials and methods

2.1. Features of the study area

The Campania region is located in the south of Italy, covering an area of about 13,600 km². The region borders the Tyrrhenian Sea and the Lazio region at the western and northern sides, respectively (Fig. 1A).

The main geological features of the Campania region are constituted by the Apennines chain which cross the areas as a backbone oriented NW-SE. The hilliest part of the chain forms the Mt. Matese in the north, Mt. Taburno and Picentini in the center, and the Mt. Alburni in the southeast. These mountains are mostly formed by sedimentary rocks such as limestone and dolostone whilst the external domains are constituted by siliceous schist and terrigenous sediments (clays, siltstone, sandstone, and conglomerate) (Bonardi et al., 1998; De Vivo et al., 2016).

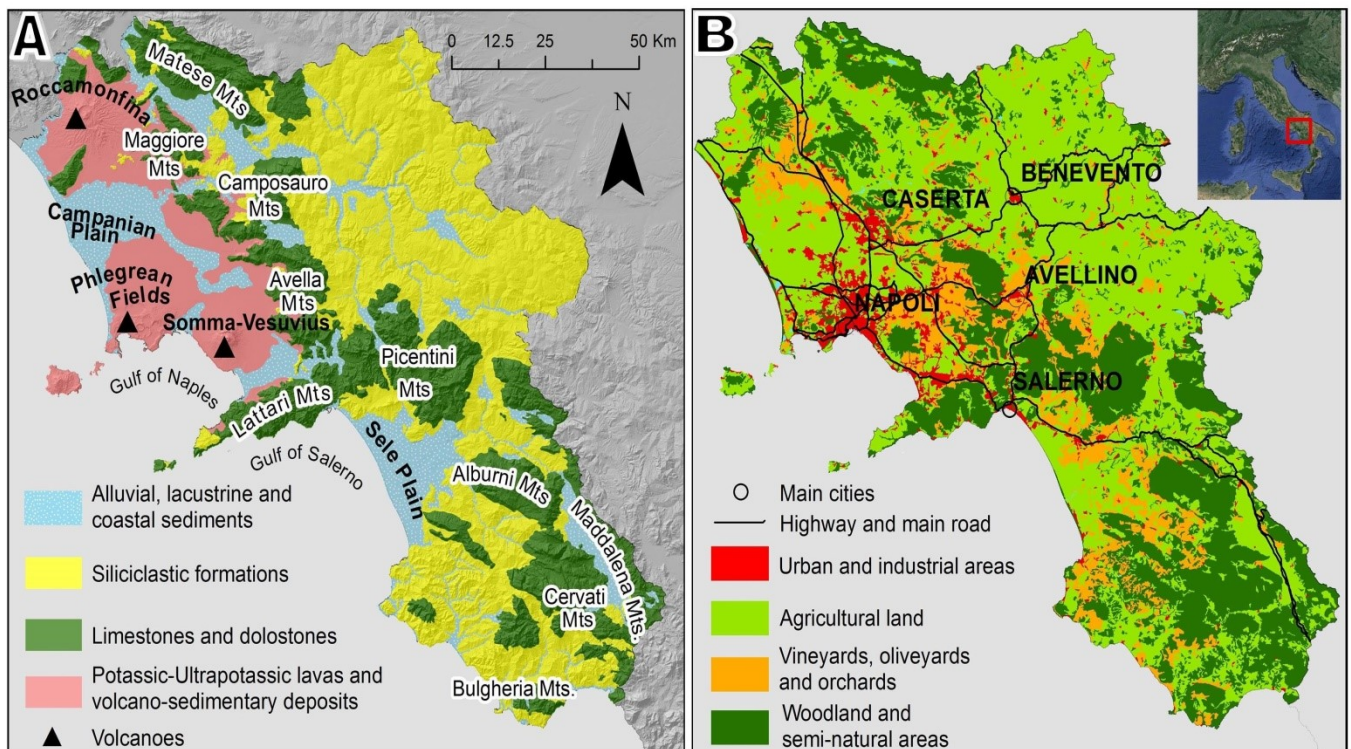


Fig. 1. Simplified geological map (A) and land use (B) of the Campania region (southern of Italy).

Igneous rocks are present in the region and are mostly formed by potassic and ultrapotassic volcanic rocks of the Roccamonfina volcano (De Vivo et al., 2001, 2010, 2016; Rolandi et al., 2003; Albanese et al., 2007) in the northwest, the Mt. Somma-Vesuvius, Campi Flegrei and Ischia volcanoes which are located nearer the coastal areas of the study region. The coastal areas and plains are constituted by alluvial, lacustrine and coastal (mixes of oceanic and terrestrial) sediments (Bonardi et al., 1998; De Vivo et al., 2016).

The hydrography of the Campania region is characterized by three main river catchments (Garigliano, Volturno and Sele) and by numerous minor streams, mostly draining towards the Tyrrhenian Sea.

The morphology of the drainage network is irregular and controlled by geological and structural features and rivers are source of water irrigation for agriculture field (Ducci and Tranfaglia, 2005).

Campania is one of the most populated regions of Italy with >5.8 million inhabitants (ISTAT, 2016). This high density of population is coupled by the presence of a large number of industrial activities, with the majority being involved in agriculture: vineyards and olive plantations – mostly in hilly areas – seasonal crops, and greenhouse products (tomatoes, potatoes, aubergines, peppers, peas, and citrus fruits) represent major resources for the region and the local economy (Albanese et al., 2007). This intensive agriculture activity occupies > 50% of the total land and occurs mostly in the coastal and mountainous areas (Fig. 1B), where fertile land and suitable soils are occurring. Unfortunately, such industrial activities are known to have a potential negative impact if not properly managed, contributing to the contamination of natural resources such as superficial and groundwater as well as soils. Campania is not immune to these problems, and some studies have already highlighted their existence and relation to its natural resources (Cicchella et al., 2005; Minolfi et al., 2018a).

2.2. Sampling procedures and analyses

From 2013 to 2015, 3669 samples were collected from topsoil of the Campania Region (13,600 km²) with a nominal density of 1 sample in each 3.2 km². Each top soil sample (from 0 to 20 cm) was made by homogenizing 5 subsamples at the corners and the centre of a 100m square, collecting approximately 1.5 kg in total. The sampling procedure followed the Geochemical Mapping of Agricultural and Grazing Land Soils (GEMAS) sampling procedure described by Reimann et al. (2014). At each sampling site, several physico-chemical parameters of the soil properties were measured, including pH, total water content, conductivity, total organic content and the geographical coordinates system recorded by geospatial positioning systems (GPS).

Chemical analyses were carried out at an international accredited Laboratory, Acme Analytical Laboratories Ltd. (now Bureau Veritas, Vancouver, Canada). The samples were analyzed after an aqua regia extraction, by a combination of inductively coupled plasma atomic emission (ICP-AES) and inductively coupled plasma mass spectrometry (ICP/MS) for “pseudototal” concentration of 53 elements (Ag, Al, As, Au, B, Ba, Be, Bi, Ca, Cd, Ce, Co, Cr, Cs, Cu, Fe, Ga, Ge, Hf, Hg, In, K, La, Li, Mg, Mn, Mo, Na, Nb, Ni, P, Pb, Pd, Pt, Rb, Re, S, Sb, Sc, Se, Sn, Sr, Ta, Te, Th, Ti, Tl, U, V, W, Y, Zn, and Zr). A sub-sample of 15 g of the sieved <2mm soil fraction was digested in 90 ml aqua regia and leached for 1 h in a 95 °C water bath. After cooling, the solution was diluted to a final volume of 300 ml using a solution of 5% HCl. The sample weight to solution volume ratio was 1 g per 20 ml. The solutions

were analyzed using a Perkin Elmer Elan 6000/9000 inductively coupled plasma emission mass spectrometer (ICP-MS). The accuracy and precision of the data was measured by comparison to known analytical standards. Calibration solutions were included at the beginning and end of each analytical run (a total of 40 solutions). Precision is $\pm 100\%$ at the detection limit, and improves to better than $\pm 10\%$ at concentrations 50 times the detection limit or greater.

2.3. Compositional data analysis

Nowadays it appears necessary to reconsider geochemical data under a compositional data analysis perspective (Aitchison, 1986; Buccianti et al., 2006, 2014, 2018; Pawlowsky-Glahn and Buccianti, 2011). A composition is defined as a sample space of the regular unit D simplex, S^D that is a vector of D positive components summing up to a given constant k , set typically equal to 1 (proportions), 100 (percentages), or 106 (ppm) by closure. It relates parts of some whole that carry relative information (ratios of components) whose sample space is the simplex (Pawlowsky-Glahn and Egozcue, 2001).

$$S^D = \{X = [x_1, x_2, \dots, x_D] \mid x_i > 0; \sum_{i=1}^D x_i = k\} \quad (1)$$

As explained in the introduction, working with log-ratio transformation such as additive log-ratio (alr), centered log-ratio (clr) and isometric log-ratio (ilr) allows to overcome some of the issues related to the complexity of geochemical data, helping to highlight the relative magnitudes and variations of the components of a composition rather than their absolute values (Buccianti and Magli, 2011). In this study, due to its orthonormal character, ilr was applied on the datasets and compared to raw and log-normal data to show how it allows normalizing the data distribution and its closure effects of geochemical data prior to statistical analyses (Figs. 2–4).

This log-ratio transformation was applied on data taking into account the compositional vectors of n parts partitioned into groups of parts presenting a certain affinity (Egozcue et al., 2003; Filzmoser and Hron, 2008, Filzmoser et al., 2009a, 2009b; Thiombane et al., 2018a). To better visualize the element distributions and possible natural and anthropogenic behaviour of the variables into the survey area, a compositional biplot was created. This is a powerful statistical tool that displays both samples and the variables of a data matrix in terms of the resulting scores and loading (Gabriel, 1971). Thus, the scores represent the structure of the compositional data into a Euclidian space based on variance and covariance matrix; moreover, they display the association structure of the dataset. The biplots present rays (or vectors) defined from the center of the plot, where their length is proportional to the amount of explained variance (communality) of the variables it represents. The interpretation of the graphic depends on the loading (rays) structures and in more details on the approximate links between rays and

samples, the distances between vertexes and their directions (Otero et al., 2005). For a full description of compositional biplots and an appreciation of their utility, several examples are available in the literature (e.g., Pison et al., 2003; Maronna et al., 2006; Filzmoser and Hron, 2008; Filzmoser et al., 2009a, 2009b; Hron et al., 2010).

When, however, biplots are used with raw data, these can be substantially influenced by the occurrence of outliers which can mislead the compositional nature of the data matrix and affect the principal components when interpreting results (Aitchison, 1986; Filzmoser et al., 2009a, 2009b). For these and others reasons log-transformed data are recommended to be used in multivariate analysis, and strengthened compositional biplots (Egozcue et al., 2003; Filzmoser et al., 2009a; Hron et al., 2010). Taking into account the singularity of the clr transformed data (Aitchison, 1986), these should be computed in orthonormal coordinates such as ilr transformed data, and back-transformed to the clr space for further interpretation. This back transformation allows preserving the linear relationship between clr and ilr coordinates (Egozcue et al., 2003). Furthermore, the application of the minimum covariance determinant (MCD) estimator (Rousseeuw and Van Driessen, 1999) allows displaying the observations to the smallest determinant of their sample covariance matrix which tend to hold the variables into a normal distribution. In this study, a classical compositional biplot (CCB) and a robust compositional biplot (RCB) were used to identify the relationships between variables using compositional raw data and log-transformed data, respectively (Fig. 5).

From the total of 53 elements analyzed for the soils, only eighteen elements were considered to test this approach, with the aim of better representing the correlation between variables and investigate more robustly their main sources in the study area. Their main descriptive statistics are shown in Table 1.

Table 1. Descriptive statistic of 3669 topsoils samples from the Campania region, Southern of Italy. RMS and Std. Deviation are the root mean square and standard deviation, respectively.

Elements	Unit	Minimum	Maximum	Mean	Median	RMS	Std. Deviation	Skewness	Kurtosis
Al	mg/kg	2100	94700	40584	41600	44014	17036	0.01	-0.75
As	mg/kg	0.6	163	12.6	12.1	14.6	7.4	6.1	87.8
Ca	mg/kg	800	295200	35495	22300	51036	36675	2.38	7.1
Co	mg/kg	0.5	79	10.7	10.3	11.7	5.1	1.6	12.3
Cu	mg/kg	2.5	2394	109.3	62.2	191	156.8	5.5	44.5
Fe	mg/kg	1600	154600	25031	2510	26282	8012	1.19	18.2
K	mg/kg	400	68200	14008	9500	18755	12472	1.39	1.2
Mg	mg/kg	700	104600	7347	5800	10483	7479	5.06	35.8
Mn	mg/kg	77	7975	863	779	970.6	443.4	5.44	55.7
Mo	mg/kg	0.06	62	1.5	1.3	2.2	1.6	18.9	637
Na	mg/kg	20	29490	3667	2600	5129	3587	1.23	1.68
Ni	mg/kg	0.5	100	16.2	14.8	19.6	10.9	2.42	9.97
P	mg/kg	50	16620	1641	1250	2063	1250	2.20	12.1
Pb	mg/kg	3.1	2052	73.8	54.2	117.7	91.7	7.7	100.2
Th	mg/kg	0.3	60	12.8	12.4	14.5	6.7	1.1	3.5
Ti	mg/kg	5	3270	1159	1240	1314	618.7	-0.23	-0.65
V	mg/kg	5	224	66.9	62	73.5	30.3	0.51	-0.29
Zn	mg/kg	11.4	3210	119	91	168.1	118.5	9.2	164.7

The number of elements was reduced to 18 variables based on 3 main criteria: 1) the removal of elements with >40% of values below the detection limit (LOD), 2) choosing arbitrary mostly two representative Rare Earth elements which are geochemically congruent and 3) choosing elements with a communality of extraction higher than 0.5 (50%) and/or common variances < 0.5 (e.g. Reimann et al., 2002). Based on the robust biplot, sequential binary partition was implemented using the same 18 variables by dividing them into specific groups of non-overlapping elements (Table 2).

Table 2. Sequential binary partition table of the 18 investigated variables and the obtained balances (Z_1 – Z_{17}). Parts coded with + and – are the elemental associations involved in the calculation of the i-th order partition, respectively.

Balances	Ti	Th	As	V	Al	P	Na	K	Mo	Cu	Zn	Pb	Ca	Mg	Fe	Mn	Co	Ni
Z_1					+	-												
Z_2							+	-										
Z_3					+	+	-	-										
Z_4	+	+	+	-	-													
Z_5				+	-													
Z_6	+	+	-															
Z_7	+	-																
Z_8						+	+	+	+	+	+	+	-	-				
Z_9									+	-								
Z_{10}											+	-						
Z_{11}										+	-							
Z_{12}								+	-									
Z_{13}													+	+	-	-	-	-
Z_{14}													+	-				
Z_{15}															+	+	-	-
Z_{16}															+	-		
Z_{17}																	+	-

Balances are particular ilr-coordinates (isometric-logratio) having orthonormal bases which can be interpreted in the D-1 (D: dimension) real space as ratios of elemental associations (Egozcue et al., 2003). Balances can be calculated using the following formula:

$$Z_i = \sqrt{\frac{rs}{r+s}} \ln \frac{(\prod_+ x_j)^{1/r}}{(\prod_- x_k)^{1/s}} \text{ for } i=1, \dots, D-1, \quad (2)$$

where the products \prod_+ and \prod_- only include parts coded with + and –, and r and s are the numbers of positive and negative signs (parts) in the i-th order partition, respectively (Egozcue and Pawlowsky-Glahn, 2005). From the established sequential binary partition and Eq. (2), Z_1 (Al/P) and Z_2 (Na/K) were calculated and ilr coordinates displayed through geospatial mapping. Balances can be interpreted considering three major cases: 1) positive balance when parts (variables) in the numerator have higher dominance with respect to parts involved in the denominator; 2) negative balance when parts involved in the numerator have lower dominance than those in the denominator (negative balance); 3) nearly zero balance when the dominance of the two groups of parts is similar. The higher or lower the positive or negative balance, respectively, the dominance of one group of parts is more pronounced.

Balances were back-transformed based on the sequential binary partition matrix and the bijection between the original space of the parts and that of the log-ratios (Egozcue et al., 2003; Egozcue and Pawlowsky-Glahn, 2005; Olea et al., 2018). The back-transformed results in the original part space for Al, P, Na and K elements concentrations were computed before applying geostatistical computations.

2.4. Interpolated and background/baseline maps

Geographical Information Systems (GIS) and technology was used to map and display data distribution and characterize the footprint, possible main sources, and the behaviour of the elements considered. For this study, one of the aims was to determine the background concentration of major elements Al, Na, K, and P in the Campania soils.

ArcGIS (ESRI, 2012) and GeoDAS (Cheng et al., 2001) were used as the main GIS tools. In particular, GeoDAS™ was used to produce interpolated geochemical maps by means of the multifractal inverse distance weighted (MIDW) algorithm (Lima et al., 2003a). In previous geochemical studies of the Campania region (De Vivo et al., 2001; Lima et al., 2003a; Cicchella et al., 2005; Albanese et al., 2007), the MIDW was chosen as an interpolation method as it preserves high frequency information (anomalies), while taking into account both spatial associations and local singularity in geochemical data (Cheng, 1999). The concentration–area (C–A) fractal method (Cheng et al., 1994) was applied to set the concentration intervals of the interpolated surfaces generated by the MIDW method, and ArcGIS™ software was used for the graphical presentation of the results (Fig. 6).

Different tools are used to determine the background and baseline concentration of elements (EPA, 2001; Reimann et al., 2005; APAT-ISS, 2006; Tarvainen et al., 2011; Cave et al., 2012; Ander et al., 2013).

They are called “traditional approaches” due to the fact that most of them do not take into account both spatial association and the data distribution local singularity. In this study, maps showing geochemical background/baseline concentrations have been obtained using the S-A (spectrum-areas) method which preserves high frequency information.

The S–A method is a fractal filtering technique, based on a Fourier spectral analysis (Cheng, 1999; Cheng et al., 2001), and is used to separate anomalies from background values starting from a geochemical interpolated concentrations map. It also uses both frequency and spatial information for geochemical map and image processing. Fourier transformation can convert geochemical values into a frequency domain in which different patterns of frequencies can be identified. The signals with certain ranges of frequencies can be converted back to the spatial domain by inverse Fourier transformation (Zuo et al., 2015; Zuo and Wang, 2016). The interpolated maps generated from geochemical data have been transformed into the frequency domain in which a spatial concentration–area fractal method has

been applied to distinguish the patterns on the basis of the power-spectrum distribution. A log–log plot (Fig. 8; Fig. 10 under plots) was used to show the relationship between the area and the power spectrum values on the Fourier transformed map of the power spectrum. The values on the log–log plot were modelled by fitting straight lines using least squares.

Distinct classes can be generated, such as lower, intermediate, and high power spectrum values approximately corresponding to baseline values, anomalies, and noise of geochemical values in the spatial domain (Fig. 9; Fig. 11, under plots), respectively.

2.5. Enrichment factor

The Enrichment Factor (EF) approach, which was historically introduced to identify the level of (economically viable) mineralisation and origin of elements in atmosphere, precipitation or seawater (Goldberg, 1972; Chester and Stoner, 1973; Peirson et al., 1974; Duce et al., 1975; Rahn, 1976; Buat-Ménard and Chesselet, 1979) was used in this study to ascertain soil contamination on a long term scale (see for example Hakanson, 1980; Sutherland et al., 2000; Abraham and Parker, 2008; Wu et al., 2011; Saeedi et al., 2012). EF is computed using the equation described below (Eq. (2)) which was first introduced by Chester and Stoner (1973):

$$EF = \frac{\left(\frac{Cx}{C_{ref}}\right)_{sample}}{\left(\frac{Cx}{C_{ref}}\right)_{background}} \quad (3)$$

where Cx is the concentration of the element under consideration and Cref is the concentration of a reference element. Here, the reference element is an element that is particularly stable in soil. In fact, the stability of the element is demonstrated by a vertical immobility and/or his chemical stability (non-degradability) (Reimann et al., 2008). Aluminium, Sc, Zr and Ti are the main elements considered in the literature to be stable, and naturally occurring in soils. In this study, the choice of the most stable element in EF computation was based on a robust statistical estimation called coefficient of variation (CV) which allows a more extensive interpretation of the variability of distribution of reference elements (Al, Zr, Sc, and Ti) using the equation:

$$CV = \frac{MAD}{MD} \times 100\% \quad (4)$$

where CV displays the variability of distribution in percentage (%), MAD corresponds to the median absolute deviation that is the median value (50th percentile) of the deviations of all concentrations from the median value of concentration and MD is the median concentration. This is a robust, nonparametric

estimate that is not affected by the presence of outliers (Reimann and de Caritat, 2005). The lower the CV value, the more the element is stable (Table 3).

Table 3. Variability Test of Al, Sc, Ti and Zr elements; MD is the Median, MAD corresponds to median absolute deviation, $MAD = \text{median} \{ |x_i - \text{median}(x_i)| \}$, that is the median value (50th percentile) of the deviations of all individual x_i values (concentrations) from the median value (concentration). CV=coefficient of variation.

Stat. Parameters	Al (ppm)	Sc (ppm)	Ti (ppm)	Zr (ppm)
MD	41600	2.3	896	5.1
MAD	10900	1.01	240	1.7
CV (%)	26.2	43.8	26.7	34.2

This study intended to investigate the most effective EF calculation by comparing the use of the reference element in continental crust (Martin and Whitfield, 1983; Peirson et al., 1974; Taylor and McLennan, 1995; Wedepohl, 1995; Loska et al., 1997) versus the use of the reference element of the local background area as advised by Reimann and De Caritat (2005a) and Sutherland et al. (2000). The variation of P EFs in the study area was then displayed by means of interpolated maps where range of EF scores were calculated based on the contamination factor in accordance with Sutherland et al. (2000) (Fig. 12, see legend).

3. Results and discussion

3.1. Univariate and multivariate analysis

Results for Na, K and P elemental distribution have been presented by combining Edaplots (top) and CP plots (bottom) with three different data type: raw data concentration (left), log-normal data (middle) and ilr transformed data (right) (Figs. 2, 3 and 4).

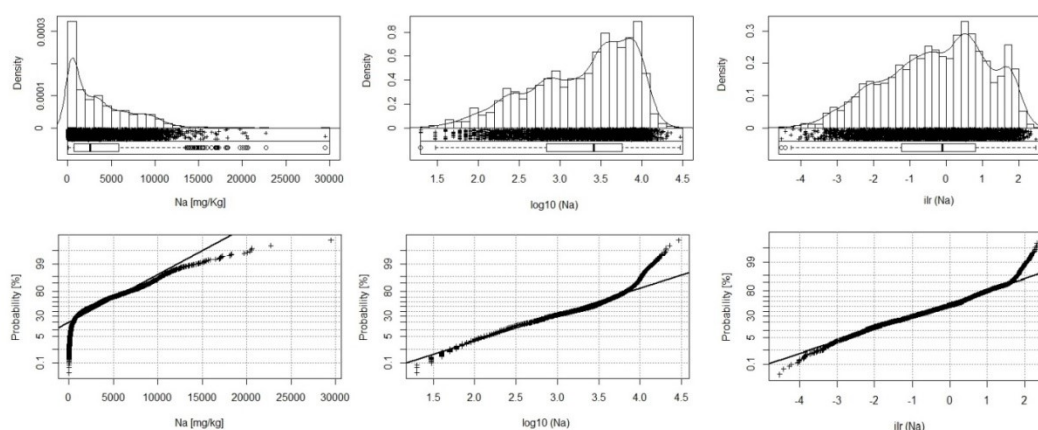


Fig. 2. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of Na of the raw data, log-transformed data, and ilr transformed data.

The Edaplots for Na (Fig. 2) show different ‘shapes’ depending on their type of data: the raw data distribution is right-skewed while the log-normal data is left-skewed. This highlights how both raw and lognormal data representation do not match well the real data distribution compared to the ilr transformation. By using the ilr transformed data, the distribution (as shown by histogram and density plot) tends to a normal data distribution. A similar result can be observed for K and P data distribution (in Figs. 3 and 4). The strength of ilr transformation in data distribution is shown in the CP plot, which displays the cumulative curve distribution where the straight-line symbolizes the most adequate model of a normal data distribution; by using the ilr transformed data, the elemental distribution fits very closely the straight line compared to raw and log-normal data, which are affected by the occurrence of outliers.

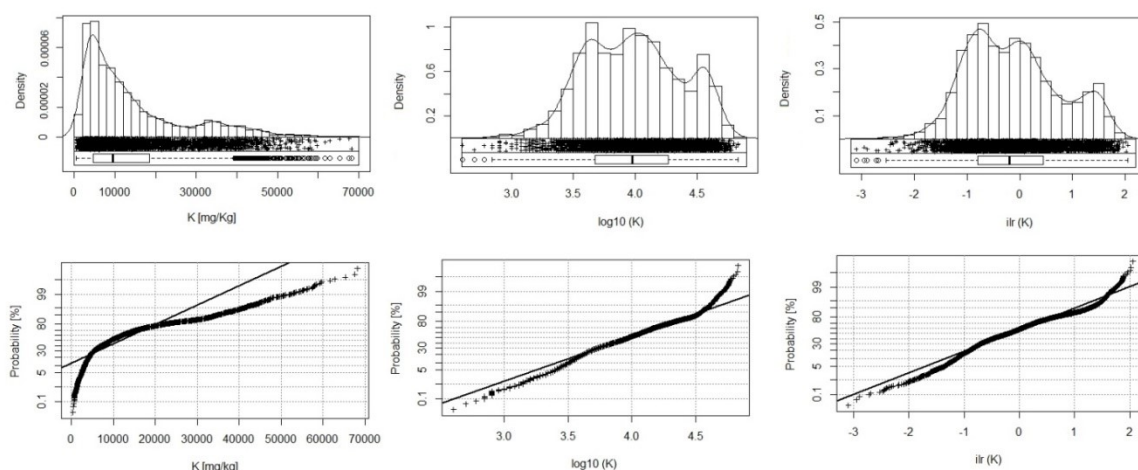


Fig. 3. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of K through the raw data, log-transformed data, and ilr transformed data.

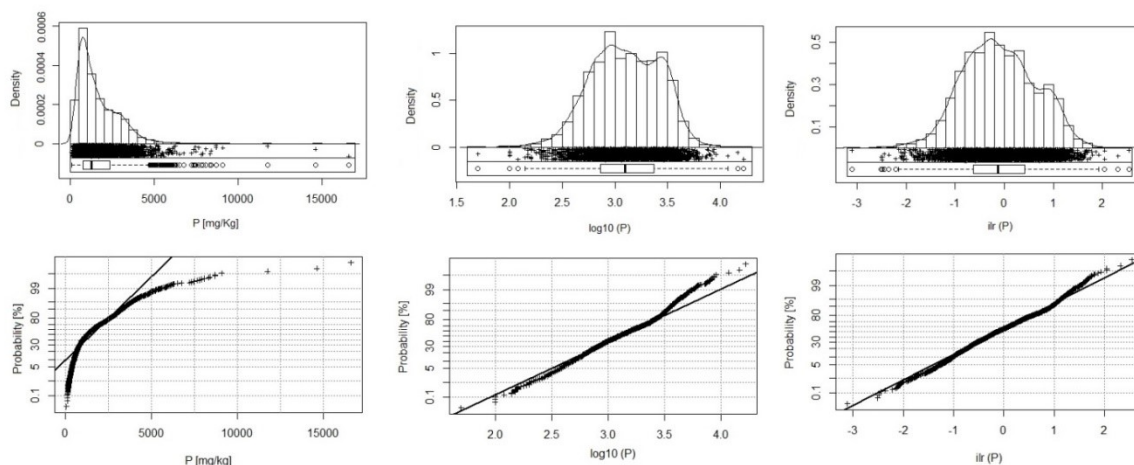


Fig. 4. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of P of the raw data, log-transformed data, and ilr transformed data.

The compositional biplot (Fig. 5), based on principal component analysis, displays the correlation and relationship between 18 analytical variables in 3669 sample points, from which the first two principal components were extracted. The principal components are presented in a compositional biplot using raw data (Fig. 5, left) and ilr coordinates clr back-transformed (Fig. 5, right).

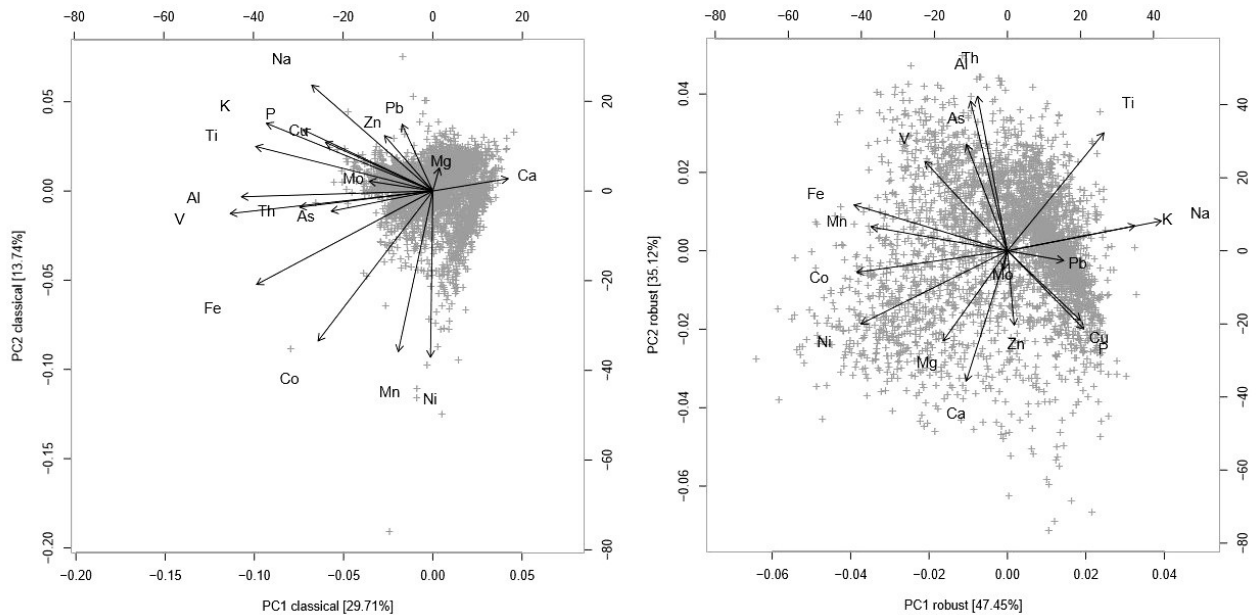


Fig. 5. Biplots for first and second principal components of factor analysis for raw data (classical, left plot) and ilr coordinates clr back-transformed (robust, right plot) of the survey area (Campania region, Southern Italy).

The total variance of initial raw data biplot (classical biplot) explains 43.45%, where the first principal component (PC1) accounts for 29.71% and the second principal component (PC2) accounts for 13.74% (Fig. 5, left). On the other hand, the robust biplot based on ilr coordinates and clr back transformed produced significantly different results (Fig. 5, right) with PC1 explaining 47.45% and PC2 explaining 35.12% of the compositional variability.

By taking into account the direction and angles formed between the vectors, and the proximity of the rays, it is possible to identify the presence of four groups of elements, which are most likely related to the geogenic features and/or the main human activities in the study areas (fig. 6).

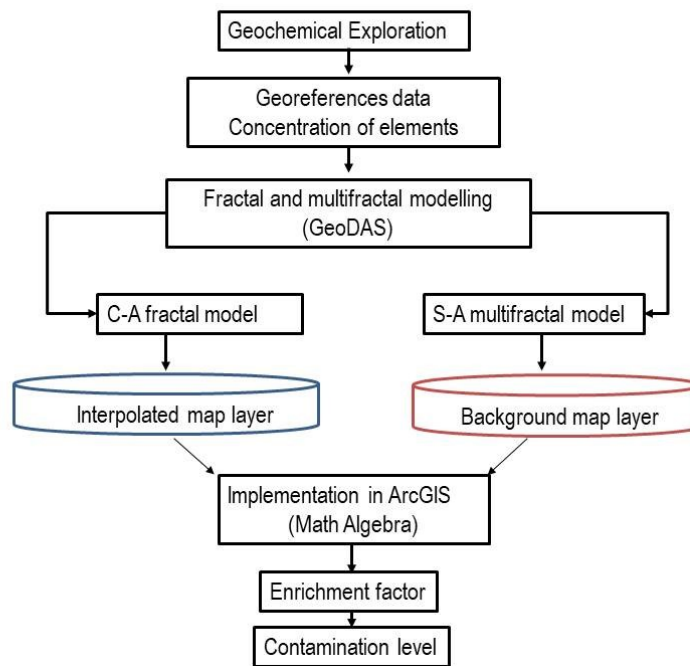


Fig. 6. Flow chart of data processing for contamination degree modelling using GIS environment.

The main groups highlighted by the biplots are:

- ✓ Fe, Mn, Co and Ni, a group of element association characterized by their rays closed to one another and tending to the same direction (classical biplot). This group of element is marked by high communalities (length of the rays) of the vectors. This may be expected given that this association is strongly related to the adsorption and coprecipitation effects operated by Fe and Mn oxides and hydroxides occurring mostly in the sedimentary deposits such as marl-sandstone, conglomerate and silico-clastic flysch deposits outcropping in the surveyed areas (Cicchella et al., 2005; Albanese et al., 2007; Buccianti et al., 2015).
- ✓ Al, Th, As, V, Ti, and Mo form an association of elements where Al, Ti and V dominate the groups with the highest length of their vertexes whilst Mo has a relatively lower communality (classical biplot). This behaviour is possibly related to the fact that these elements are mostly immobile during weathering phenomena of the parental rocks and mostly remaining in the residual fraction of soils. This group could therefore be directly related to the parental geology of the surveyed areas which are dominated by the influence of pyroclastic deposits from different eruptions of nearby volcanoes such as Roccamonfina, Vesuvius, Campi Flegrei (De Vivo et al., 2010).
- ✓ Na, K, P, Cu, Pb and Zn elemental association is dominated by a high communality of Na and P as well as the vicinity of their rays (classical biplot). This elemental association probably

reflects the potassic and ultrapotassic rock formations that occur throughout the majority of the slope of Naples and Benevento areas, associated to the lava and pyroclastic volcanic activity of Mt. Somma–Vesuvius and Roccamonfina (Lima et al., 2003b; Albanese et al., 2013). Zinc and Pb display short vectors which are poorly characterized and seem to be only partially correlated to the others element of this association. These two elements may be related to anthropogenic activities such fossil fuel combustion, as well as industrial and vehicular emissions release (Cicchella et al., 2005; De Vivo et al., 2016).

- ✓ Ca and Mg appear to be correlated and present both a short length of their rays (classical biplot). The communality of Ca is larger and seems to be independent of all others elements due to the fact that the angles formed are $>90^\circ$ compared to Mg. This confirms the high correlation between Ca and Mg, which might be possibly related to the limestones and dolostones of the Mt. Picentini, Mt. Lattari and Mt. Cervati. On a closer observation, the elemental association Na, K, P, Cu, Pb and Zn could be reduced to two main subgroups, where the variables K and Na are strongly overlapping on to each other (robust biplot). This highlights the high correlation between these two elements occurring mostly in potassic and ultrapotassic rock formations throughout the surveyed areas. Phosphorous and Cu are also highly correlated both with high communalities, where P seems to be independent of the Na/K vertexes, forming almost an angle of 90° (robust biplot). Interestingly, Cu seems to be poorly correlated to a geogenic origin, as the direction of its ray compared to those of the group of sulphide elements (e.g. Co and Ni) have an angle up to 90° . This may signify that Cu is independent from other naturally occurring sulphide elements, whilst it seems to be highly correlated to P in most soil of the surveyed area. One potential explanation is that P and Cu may be related to agriculture activities, with large areas cultivated as vineyards, where the use of pesticides and phosphate fertilizers is very high (Cicchella et al., 2005). For the Al, Th, As, V, Ti, and Mo association, it is observed the dissociation of Ti variable with a high communality of the ray (robust biplot). Titanium is considered as an immobile and stable element due to low mobility and is mostly found in volcanic materials (Egli et al., 2008).

3.2. Geochemical elemental distribution in the survey areas

Based on the robust biplot, 18 elements have been chosen to perform sequential binary partition and obtain balances (specific ilr-coordinates) (Table 2). In this section, Balances Z1 (Al/P) and Z2 (Na/K) will be displayed to reveal the data for the elements of main interest (Al, P, Na and K).

The first balance Z1 (Al/P) map, ranging from 2.89 to 4.79 reveals a higher proportion of Al in correspondence to large volcanic complexes like Mt. Roccamonfina and Phlegraean fields (Fig. 7A). In

addition, high proportion of Al is also highlighted in correspondence of part of the Lattari range, along the Apennines and in patches at the southern part of our study area. Scheib et al. (2014) highlighted that the Mt. Roccamonfina volcano is characterized by pyroclastics rocks with high level of elements such Al, Th and Ti, as well as in the Campanian Ignimbrites in the Apennines (De Vivo et al., 2010).

In contrast, the higher proportions of P (Fig. 7A) in correspondence to lower values of coordinate (ranging from 0.6 to 1.95) are found around Mt. Somma-Vesuvius, and in several areas of the eastern region of our study area, where large agricultural fields (e.g., vineyards and orchards) are located.

The second balance map Z2 (Na/K) (ranging from -0.65 to 0.41) shows the dominance of Na and K in the study area (Fig. 7B).

In fact, higher proportion of Na can be observed in correspondence to Mt. Roccamonfina, Phlegrean fields, and Ischia Island. Sodium may be related to the potassic and ultrapotassic rocks and volcano-sedimentary deposits from major sector collapse of volcanoes in the study area (Scheib et al., 2014; De Vivo et al., 2016). In contrast, the higher abundance of K corresponding to lower balances (ranging from -3.26 to -2.48), can be observed in the southern part of our study area.

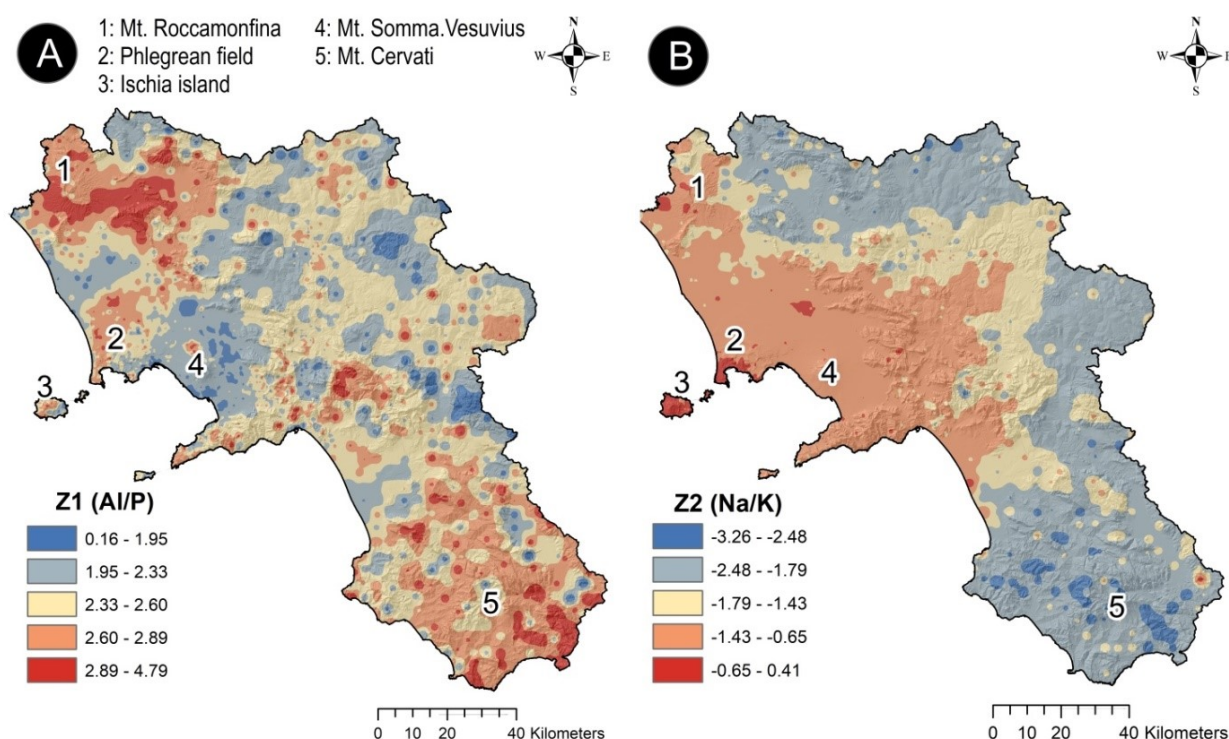


Fig. 7. (A) The interpolated Z1 map. Note the red and blue colours highlight higher and lower proportion of Al and P, respectively. (B) The interpolated Z2 map. The red and blue colours highlight higher and lower proportion of Na and K, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In fact, Thiombane et al. (2018a) showed a high enrichment of K in silicoclastic deposits dominated by flysch series in southern part of our study area. Furthermore, pyroclastic rocks from different eruptions of nearby volcanoes (Roccamonfina, Vesuvius, Phlegrean Fields - De Vivo et al., 2010; Buccianti et al., 2015; Mt. Vulture and Aeolian islands - Peccerillo, 2005; Scheib et al., 2014) are found in this area. The back-transformation of balances based on the partition matrix helped to display results in the original part space giving the same values of Al, P, Na and P elements concentration because of the bijection between the original space of the parts and that of the *ilr*-transformation (Egozcue et al., 2003; Olea et al., 2018).

The concentration of Al, ranging from 2344 to 94,334 mg/kg with a mean value of 32,918 mg/kg, was separated into five ranges according to C-A fractal plot (Fig. 8A, plot below). The lowest concentration values roughly ranging from 2344 to 30,000 mg/kg, are found in the north-eastern and south-western part of the study area in correspondence with the Apennine chain and the Cervati Mt., respectively. The highest concentrations (up to 57,000 mg/kg) are found in soils on the slope of the volcanoes (Mt. Somma-Vesuvius and Roccamonfina), surrounding the Mt. Matese and the Mt. Lattari (Fig. 8A). The elevated concentration of Al in soils surrounding the volcanoes is possibly related to the parental pyroclastics which subsequently formed soils (De Vivo et al., 2016). On the other hand, in the Mt. Matese and Mts. Lattari, Al concentrations could result from the occurrence of several imbrications of bauxite minerals which were exploited in the first part of the 20th century (Mondillo et al., 2011).

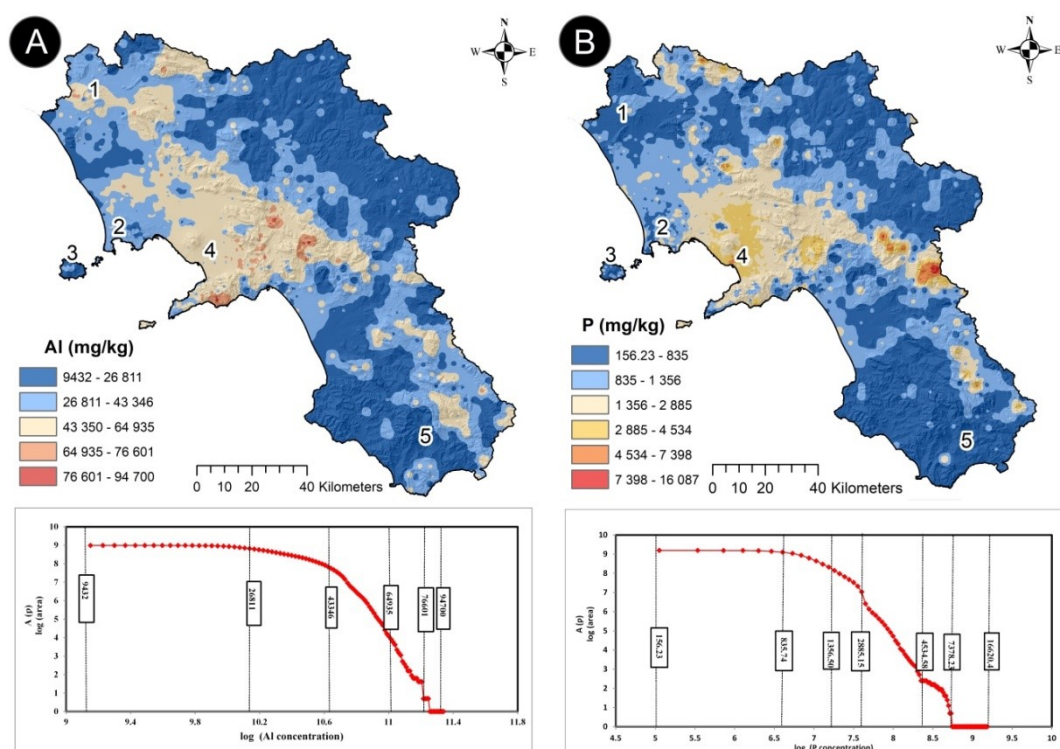


Fig. 8. Interpolated maps of Al (A) and P (B) elemental distribution in the survey area; ranges of concentration are based on the C-A fractal plot held below.

The background/baseline map of Al (Fig. 9A) shows the highest concentration ranging from 51,000 to 61,000 mg/kg in correspondence with soils around the Roccamonfina and Vesuvius volcanoes. Phosphorus content in Campania soils ranged from 156 to 16,087 mg/kg (Fig. 8B) with a mean value of 1011 mg/kg, which corresponds to the mean level of the European soil (Tóth et al., 2014).

Average concentrations (ranging from 64,935 to 75,000 mg/kg) increased significantly near Mt. Somma-Vesuvius and the highest values (up to 75,000 mg/kg) were found in Vitulano municipality (Benevento Province), and in the nearby Lioni and Laviano districts (Avellino Province). In particular, the corresponding underlying geology does not seem to be able to account for this P elemental anomaly: the soils of Vitulano district are mostly from sandstone, flysch deposits and limestone whereas Lioni and Laviano districts soils are mostly from limestones.

The high P concentration might be possibly related to anthropogenic activities related to the large fertilizers use in agriculture. Part of the Vitulano municipality lies in fact in an area with intensive vineyard occurrence. The background/baseline P distribution (Fig. 9B) is characterized by >95% of the survey area with P values <1500 mg/kg. This range of concentration is usually characteristic of siliciclastic, limestone and dolostone deposits/geology. Greater concentrations ranging from 2200 to 4600 mg/kg were found on the slopes of the Mt. Somma-Vesuvius, and Lioni and Laviano districts, possibly linked to geogenic and anthropogenic sources, respectively.

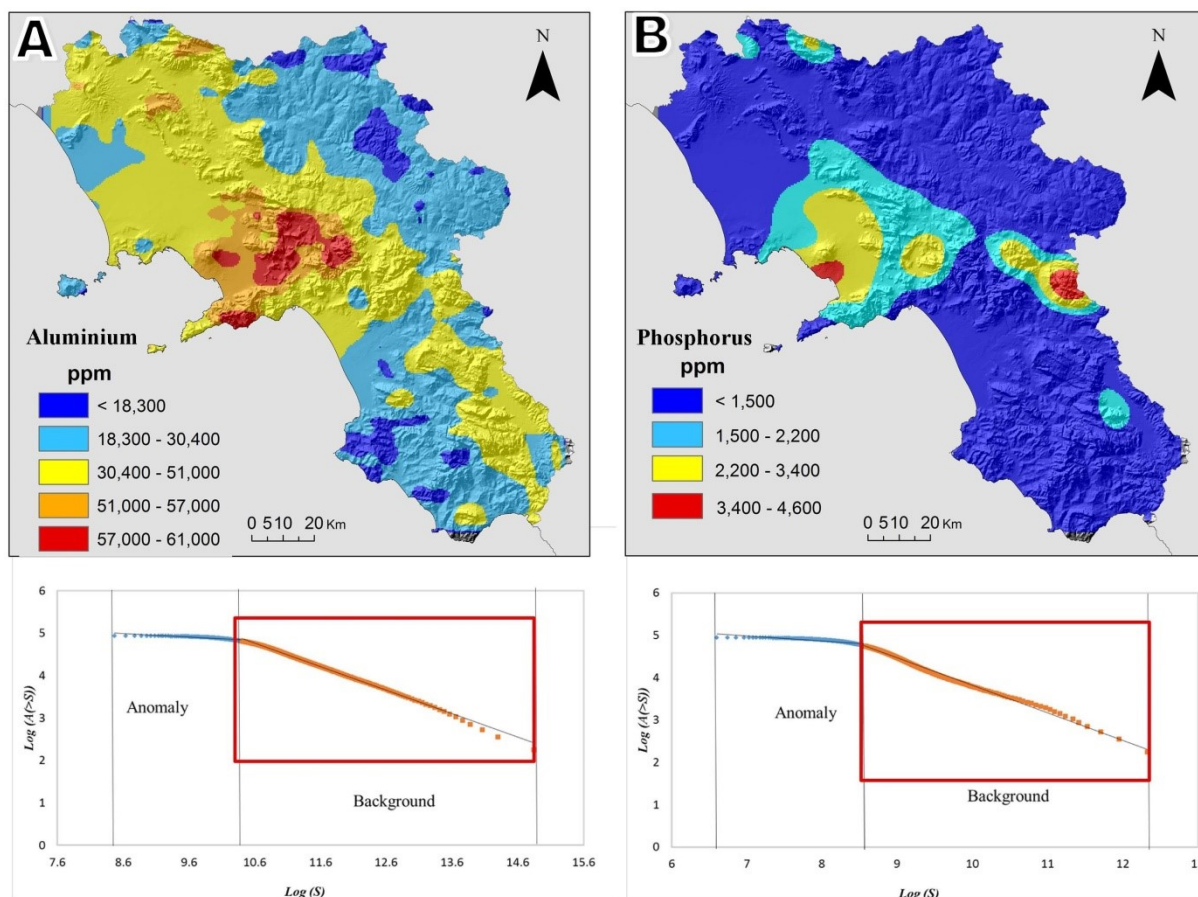


Fig. 9. Background/baseline maps of Al (A) and P (B) elemental concentration in the survey area.

Sodium and K concentrations ranged from 20 to 17,592 mg/kg with a mean concentration of 1265 mg/kg, and from 804 to 63,850 mg/kg with a mean concentration of 6852 mg/kg, respectively (Table 1). The highest values of Na (up to 10,900 mg/kg) were found in soil samples on the slopes of the Somma-Vesuvius and Roccamonfina volcanoes, in the Phlegrean Fields and Ischia Island whereas those for K (up to 41,522 mg/kg) were found mostly on soils surrounding the Mt. Somma- Vesuvius (Fig. 10A and B). These values could reliably be attributed to the occurrence of volcanic rocks and related soils in the surveyed areas.

These formations are dominated by potassic and ultrapotassic lavas and pyroclastic materials. The interpolated maps reflect hence the concentrations of Na and K elements in such rocks and pyroclastics formations linked to Quaternary volcanic activities (Peccherillo, 2005; Albanese et al., 2013; Buccianti et al., 2015). In general, the soils of the Vesuvian area, formed from a more recent volcanic activity (Joron et al., 1987; De Vivo et al., 2003; Lima et al., 2003b) are much richer in Na and K than the soils on the much older Roccamonfina volcano. This is due to the fact that Na and K are relatively mobile and easily leached elements in the surficial environment.

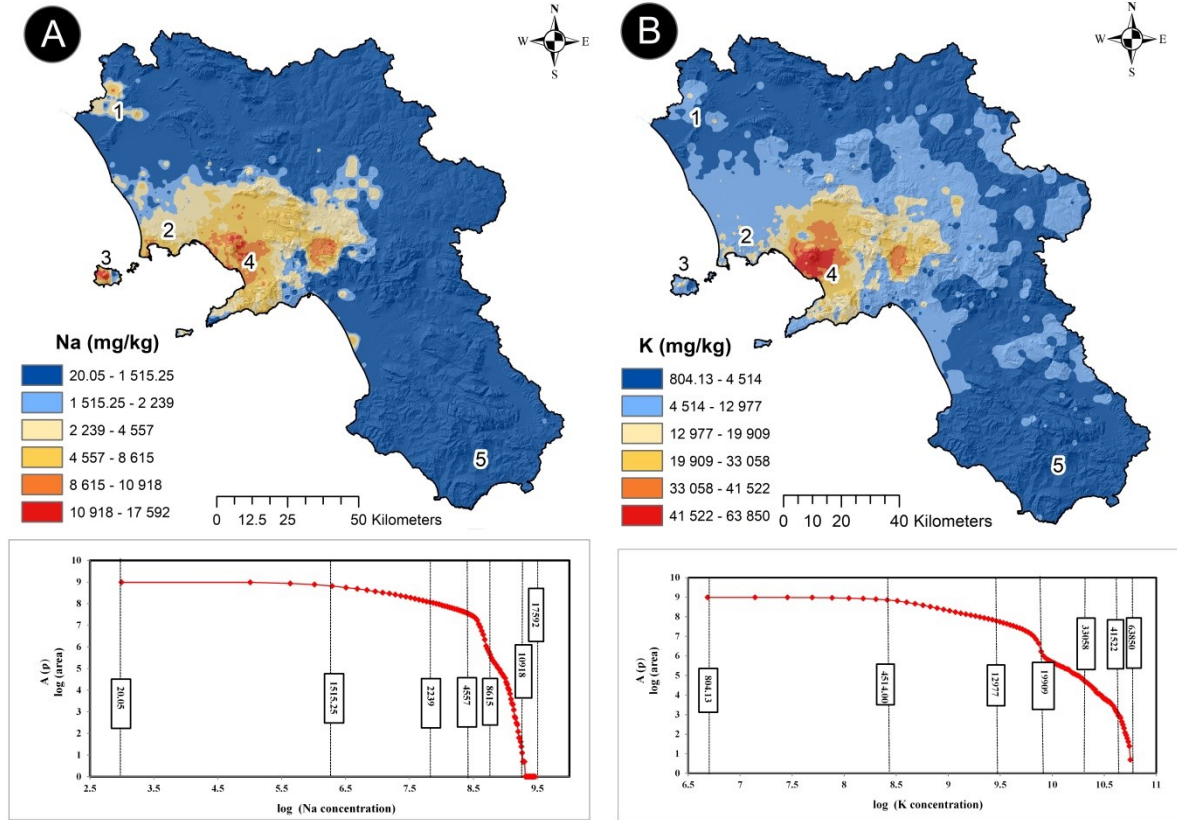


Fig. 10. Interpolated maps of Na (A) and K (B) elemental distribution in the survey area; ranges of concentration are based on the C-A fractal plot held bellow.

The background maps of the Na and K (Fig. 11A and B) show low concentration ranging from 20.1 to 3200 mg/kg and from 804 to 9100 mg/kg, respectively. These relatively low ranges of concentration of Na and K were found to correspond to the same lithologies. At regional level, the highest concentration of Na and K was found in soil on the slopes of the Roccamonfina and Mt. Somma-Vesuvius, ranging from 3200 to 9700 mg/kg and from 9100 to 46,900 mg/kg, respectively. This allows delimiting two ranges of background concentration of Na and K related to the local geology (Table 4).

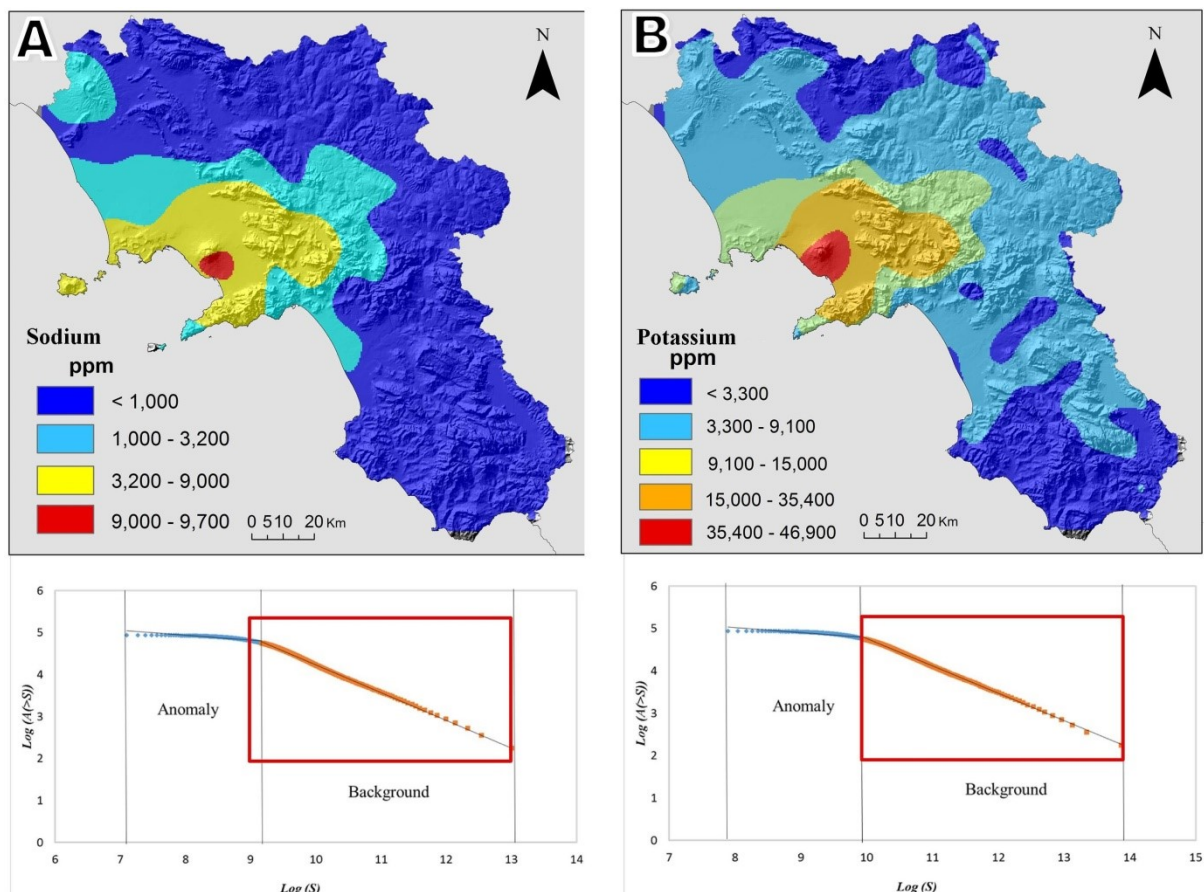


Fig. 11. Background/baseline maps of Na (A) and K (B) elemental concentration in the survey area.

3.3. Enrichment factor or phosphorus degree of contamination

The variability coefficient (CV) of the reference elements was measured as: 26.2% (Al), 43.8% (Sc), 26.7% (Ti) and 34.2% (Zr - Table 3). Aluminium displayed the lowest value of CV, followed closely by Ti, Zr and Sc. Al and Ti are mostly related to processes forming and presence of oxides (Al_2O_3 and TiO_2) in clastic materials and are not easily affected by weathering processes. These results prove that Al remain the most stable element and was therefore chosen as reference variable to determine the P enrichment factor for this study too.

The enrichment factor scores were calculated for P using both the Al elemental concentration in continental crust (Fig. 12A), and the local Al background concentration in the survey areas (Fig. 12B).

The reference with the continental crust value (Fig. 12A) showed a lower EF score ($EF < 2$) in the northern and southern parts of the region, with medium EF scores (ranging from 2 to 4) comprising >50% of the study area. Higher EF scores (ranging from 20 to 40), corresponding to anomalous enrichments, were found in soil on the slopes of the Mt. Somma-Vesuvius and in the Lioni and Laviano districts. Vesuvian areas (Fig. 9) displayed higher concentration of P which can though be related to the underlying background concentration, where the enrichment can be explained by geogenic (volcanic)

source. Lioni and Laviano districts lie in areas characterized by limestone-derived soils; a high P enrichment factor could therefore be linked to anthropogenic activities. The highest range of P enrichment factor scores (> 40) is registered in the provincial territory of Benevento (particularly towards the Vitulano municipality). Soils in this area are derived from sandstone, flysch deposits and limestone, where intensive agriculture activities, such as vineyards and olive plantations, are practiced.

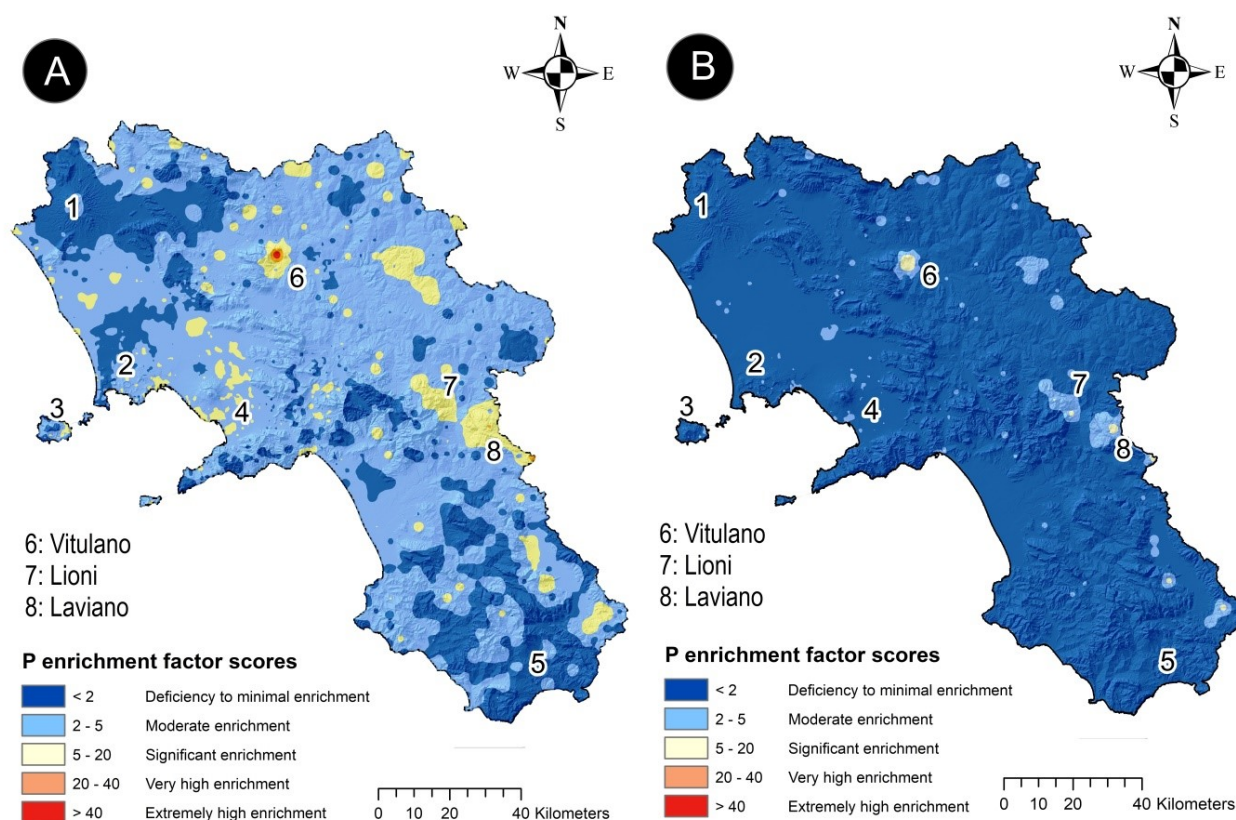


Fig. 12. Interpolated maps of the enrichment factor scores into the study areas; (A) map created using the continental crust reference values of Al and P (Wedepohl, 1995); (B) map is based on the local reference background concentrations of Al and P (this study) as reference elements in each pixel.

The extremely high P enrichment may be related to the use of phosphate fertilizers in agriculture practises in this area.

The calculation of P enrichment factors with local background concentrations (Fig. 12B) presented different distribution and intensity of EF scores in the study area. The Vesuvian areas presenting previously significant enrichment (EF ranging from 5 to 20, Fig. 12A) were now presenting a minimum enrichment instead (EF < 2 , Fig. 12B). The Vesuvian area are characterized by a high P background concentration related to the underlying parental volcanic rocks which represent materials for the subsequently formed soils throughout this area. Furthermore, Al was found at lower concentration values compared to its value in the continental crust (Table 4). Indeed, EF calculation using the P and Al local background concentrations provides a smoother EF in relation to the Vesuvian area compared to

their actual local variabilities. A similar result was observed in the area surrounding Lioni, Laviano and Vitulano municipalities. EF calculated in the classic way (Fig. 12A) showed a very high (EF ranging 20 to 40) and extremely high enrichment (EF > 40) as opposed to the moderate (EF ranging from 2 to 5) and significant enrichment (EF ranging from 5 to 20) shown by EF calculated using the local background (Fig. 12B). Using the local background concentration of Al and P (Fig. 12B), the widespread enrichment created by calculation made using the continental crust values as a reference fades away, replaced by a much narrower enrichment that take into account the local variability of P and Al, and highlights the anthropogenic P inputs observed around the Lioni and Laviano districts, and even more significantly in the Vitulano municipality where it may be related to the use of phosphate fertilizers in agriculture activities. A continental crust concentration based EF 'hides' the anthropogenic input behind a more general enrichment due to the fact that it does not consider the local geological variability (Reimann and de Caritat, 2005; Albanese et al., 2013; Zuzolo et al., 2017). However, when taking into account the local concentration, the anthropogenic input can be more clearly distinguished from the geogenic enrichment factor. In other words, by using as a reference the local background, it is easier to isolate anthropogenic enrichment factors from the geogenic ones.

Table 4. Geogenic background/Baseline value ranges for Al, P, Na and K elemental concentration in the continental crust (Wedepohl, 1995) and in soils of the Campania region according to the lithology.

Elements	Wedepohl, 1995*	Local background in this Study		
	Continental Crust	Siliciclastic deposits	Limestone and Dolostone	Volcano-clastic deposits
Al (mg/kg)	150500	9432 - 30400	30400 - 51000	51000 - 61000
Na (mg/kg)	35600	20.4 - 1000	20.4 - 1000	1000 - 9700
K (mg/kg)	31900	804 - 9100	804 - 9100	9100 - 46900
P (mg/kg)	1500	156 - 1500	156 - 1500	1500 - 4600

4. Conclusions

Evidence from this study showed that compositional data transformations such as ilr transformation can help to solve the outlier artefacts and moves the composition sample space to the Euclidean Real space R^{-1} , which is failed by the classic statistical data transformation (Filzmoser and Hron, 2008).

The multivariate and integrated approach applied in this study on a multi-elemental geochemical dataset allowed to highlight the correlation between variables and helped identifying the main sources of elements in the surveyed region. Biplot based on transformed data were able to highlight the main geological features as well as the potential input of anthropogenic activities in the Campania region. In particular, four association complexes were identified:

- 1) Fe, Mn, Co and Ni: associated to the coprecipitation of Mn and Fe oxides-hydroxides in flysch and arenaceous material;
- 2) Ca and Mg: associated to limestones and dolostone outcrops occurrence of the Mt. Picentini, Mt. Lattari and Mt. Cervati limestones;
- 3) Al, Th, As, V, Ti, and Mo: associated to the pyroclastic coverings;
- 4) Cu and P: potentially associated to the agriculture activities through use of phosphate fertilizers.

The robust biplots allowed to display elements in a wider space and provided grounds for an enhanced data interpretation: the length of the rays, which is linked to the variability of ilr data clr back transformed as opposed to the variables themselves, permitted to emphasize their potential sources; the groups of elements highlighted by proximal rays, were used as evidence of either local geological features and/or anthropogenic activities.

In addition, the interpolated maps by use of C-A fractal plot helped to distinguish element distributions related to their main sources, whereas using S-A multifractal plot allowed to display the background concentration Al, Na, K and P elements as reference values for Campania soils.

The maps of P Enrichment Factor scores using as reference the continental crust and the local background values showed:

- 1) A significant P enrichment (from 5 to 20) in soils on the slopes of the Mt. Somma-Vesuvius related to the underlying parental volcanic rocks (geology), where the geogenic component represent clearly the natural background. This enrichment disappeared when the local background reference of Al and P are used. In fact, due to the geogenic source of the P and Al in these areas, the EF tended to decrease when using the local background, confirming that those areas were not contaminated by human activities.
- 2) The highest range of P EF factor scores (> 40) were registered in the Benevento provincial territory (particularly in Vitulano district), Lioni and Laviano districts, with values decreasing two times when using as reference the local background values. Taking into account that the underling geology of these areas could not influence these high P EFs, it is very likely that these values may have an anthropogenic source. The above provincial districts fall in a territory where intense agriculture activities are present, allowing inferring that such soils may be affected by use of phosphate fertilizers.

As a general observation, the findings from this study confirm the validity of using local background concentrations to better identify the 'real' degree of contamination as opposed to generalised continental crust values, which could emphasize 'spurious' enrichment due mainly to natural local concentrations of elements. From an applied point of view, the integrated approach applied here provided a more robust qualitative and quantitative evaluation, highlighting new and vital information on the distribution and patterns of key elements (Na, K, Al and P) in soils of the Campania region. The findings from this

investigation strongly point towards highly desirable follow up studies in clearly identified and discrete areas displaying high P EFs: this would allow a more detailed and thorough assessment of P footprints, which could then be used for a comprehensive human health risk evaluation due to direct and indirect exposure. This is particularly important, when considering that, if P was linked to fertilizers source, these could include impurities such as Cu, As, Zn and Pb, well-known potentially toxic elements (PTEs).

Section 2.2

Soil contamination compositional index: A new approach to quantify contamination demonstrated by assessing compositional source patterns of potentially toxic elements in the Campania Region (Italy)

This section has been published in:
Journal of Applied Geochemistry
Volume 96, September 2018, Pages 264-276

Soil Contamination Compositional Index: a new approach to quantify contamination demonstrated by assessing compositional source patterns of potentially toxic elements in the Campania Region (Italy)

Abstract

Potentially toxic elements (PTEs) are a major worldwide threat to the environment due to the constant global increase in industrial activity and urbanisation. Several studies have provided detailed maps and a better understanding of the spatial distribution patterns of PTEs in different matrices, but the majority of these studies have simply neglected the compositional nature of geochemical data. The aims of this study are to reveal the compositional behaviour and relative structure of 15 PTEs (subcomposition) in Campania, one of the most contaminated regions in Italy, and to quantify the spatial abundance and identify the possible origins of these PTEs. Robust compositional biplots were used to understand the natural grouping and origin of the PTEs. Ratios of specific subcompositions (balances) of PTEs were calculated to map the spatial patterns and identify the spatial variability of the PTEs. This study presents the preliminary steps needed to quantify and analyse the relative difference in the spatial abundance of PTEs by applying a compositional abundance index. In addition, a new soil contamination compositional index (SCCI) was elaborated to quantify topsoil contamination by the 15 PTEs and related subgroups following the compositional structure of the geochemical data. The elevated spatial abundance of the 15 PTEs is related to highly urbanised (Naples and Salerno), highly industrialised (Solofra) and intensely cultivated areas (Sarno River Basin), where the high dominance of elements from the anthropogenic subgroup (Pb, Sb, Sn and Zn) and high SCCI values suggest that contamination is from anthropogenic sources. The high spatial dominance of elements from the volcanic rock subgroup (As, Be, Se, Tl and V) in these same areas is likely related to geogenic sources, including alkalic pyroclastic rocks. Although the high spatial abundance of Group B elements (Cd, Cr, Co and Ni) is related to Terra Rossa soils and shaley facies of siliciclastic rocks of the southern Apennines, these same elements can also reach high abundances and reflect contamination (i.e. high SCCI values) from urbanised and industrialised areas due to e.g., tanneries and alloy production. Other high spatial abundances of the 15 PTEs with little or no contamination (i.e. very low SCCI values) can be related to nearby carbonate massifs, where a mixture of geogenic factors including weathering, advanced pedogenic processes, adsorption and co-precipitation with Fe-/Mn-oxy-hydroxides and the presence of pyroclastic material might all be responsible for an increase in abundance. The lowest spatial dominance of the 15 PTEs occurs in the north-eastern and southwestern siliciclastic zones of the Campania Region, where there is a low level of urbanisation and industrialisation and therefore contamination from any source can be excluded.

1. Introduction

One of the main objectives of environmental geochemistry research is to constrain the spatial distribution patterns of various organic and inorganic elements and compounds to determine whether sources are geogenic and/or anthropogenic. Geochemical mapping and the delimitation of mineralisation or contamination/pollution areas has always been the focus of exploration and environmental geochemistry. Several graphical techniques including proportional dots (Björklund and Gustavsson, 1987), statistical methods like Mean + 2 Standard Deviation (Hawkes and Webb, 1962; Reimann and Garrett, 2005) and Median + 2 Median Absolute Deviation (Reimann et al., 2005), cumulative probability plots (Tennant and White, 1959; Lepeltier, 1969; Sinclair, 1974), upper-fence criteria in Tukey's boxplots (Tukey, 1977) and stated percentiles (e.g., 95th, 98th) in box-and-whiskers plots (Kürzl, 1988) have all been used to define geochemical threshold values and outliers. The separation of elemental anomalies from baseline and background values using various fractal methods such as the Concentration-Area (C-A) or Spectrum-Area (S-A) techniques and local singularity analysis have been successfully applied (Cheng, 1999; Cheng et al., 1994, 2000; Lima et al., 2003a, 2005; Albanese et al., 2007, 2015; Zuo et al., 2015; Zuo and Wang, 2016; Parsa et al., 2017; Minolfi et al., 2018a; Rezza et al., 2018). Potentially Toxic Elements (PTEs) have always been given special emphasis because the accumulation of these elements in different matrices can cause soil and land degradation that can then be transferred to the human body from dermal contact, inhalation or ingestion through the food chain and drinking water (Lim et al., 2008; Ji et al., 2008; Varrica et al., 2014). PTEs are a major environmental concern because concentrations are constantly rising due to accelerated population growth and therefore increased urbanisation and industrialisation that generate a wide range of anthropogenic contamination sources (Albanese et al., 2010; Guillén et al., 2011; Wang et al., 2012; Wu et al., 2015). PTE compounds are generally non-biodegradable with long biological half-lives that tend to accumulate in soils by adsorption to clay minerals and organic matter (Kabata-Pendias, 2011). However, PTE bioavailability is influenced by different physicochemical processes (e.g., pH, Eh) and physiological adaptations (Yang et al., 2004; Skordas and Kelepertsis, 2005; Barkouch et al., 2007; Kabata-Pendias, 2011; Zhao et al., 2013, 2014). Several single and integrated contamination indices have been used, such as the Enrichment factors (Chester and Stoner, 1973), Geoaccumulation Index (Muller, 1969) and Single Pollution Index (Hakanson, 1980; Muller, 1981), to quantify the contamination status of different environmental media. Indices using background/baseline values (e.g., Single Pollution Index) for reference are straightforward, but are not scale-invariant meaning that the change in units of the concentrations modifies the results of the analysis (Aitchison and Egozcue, 2005; Pawlowsky-Glahn et al., 2015a). The critical reviews of element ratio variations and Enrichment factors (EFs) were made by Reimann and de Caritat (2000,2005) claiming that their values vary and are dependent on the

different parent rock materials and chosen reference media as well as reference elements. In addition, these indices do not take into account the different biogeochemical processes, the natural fractionation of elements or differential solubility of minerals which may have remarkable impact on elemental enrichment/contamination (Reimann and de Caritat, 2000, 2005b). Sucharovà et al. (2012) emphasised that high EFs or top-/bottom soil ratios reflect the geochemical decoupling of the lithosphere from the biosphere rather than contamination. They concluded that regional distribution of the raw data from multi-medium provides a better indication of the extent and impact of pollution than ratios calculated based on misconceptions. Fabian et al. (2017) introduced a new method for detecting and quantifying diffuse contamination based on the analysis of cumulative distribution functions. This method revealed local contamination sources and efficiently monitored diffuse contamination at the continental to regional scale. However, geochemical data are inherently compositional data; therefore indices measuring the contamination/pollution should be considered the main principles of compositional data analysis.

This study aims to better understand the compositional behaviour and relative structure of a subcomposition of 15 elements, which are listed as PTEs for the Campania Region (D. Lgs. 152/2006), and to quantify the spatial abundance and identify the possible origins of PTEs. We put special emphasis on the ratios of specific subcompositions (balances) of PTEs by mapping the spatial patterns and making profiles to reveal the spatial variability. This study presents the preliminary steps to analyse the spatial abundance of PTEs by applying a compositional abundance index. A new soil contamination compositional index was elaborated to quantify the topsoil contamination from the 15 PTEs and associated subgroups following the compositional structure of geochemical data.

2. Study area

2.1. Geology and Landuse

The main geology and Landuse of the Campania region have been described in the section 2.1.

3. Materials and methods

3.1. Sampling procedure

Sampling design and procedure follow Geochemical Mapping of Agricultural and Grazing Land Soils (GEMAS) sampling protocol described in the section 2.1.

3.2. Methods

3.2.1. Statistical analysis

A wide range of summary statistics [lower and upper quartiles (Q1, Q3), maximum, minimum, mean, robust coefficients of variation (CVR), median absolute deviation (MAD), kurtosis and skewness] were

applied in this study using log-transformed data that was then back-transformed to describe the central tendency and variability of the investigated PTEs. Although the log-ratio transformation of data is more relevant in compositional data analysis, the summary statistics output expressed in the raw concentrations of single elements is also meaningful and more easily interpretable.

Tukey boxplots were also generated with log-scaling to visualize the distributions and distinguish the univariate outliers of each of the 15 PTEs. The intervention limits of the 15 PTEs for residential and industrial areas established by Italian law (D. Lgs. 152/2006) were also displayed. To serve as an additional reference, the median values of the 15 PTEs from European agricultural soils (Reimann et al., 2014) were also included and the ratio between the Campanian topsoil medians and European counterparts was calculated.

3.2.2. Compositional multivariate analysis

Geochemical data are typically compositional in nature with positive vectors whose elements (parts) describe quantitatively relative contributions to a whole (Aitchison, 1986; Pawlowsky-Glahn and Buccianti, 2011; Pawlowsky-Glahn et al., 2015b). Egozcue et al. (2003) introduced the isometric log-ratio (ilr) transformation to generate coordinates with an orthonormal basis, which cannot be affected by intrinsic distortion of scatterplot matrices and are subcompositionally coherent. One way to get ilr coordinates is fully evinced in the section 2.1. The robust biplot was created by considering not only the 15 PTEs but also 9 major elements (Al, Ca, Fe, K, Mg, Mn, Na, P and Ti). The major elements were included to better interpret and delineate the subgroups of PTEs and distinguish geogenic factors from anthropogenic ones. However, for the construction of balances, the focus was only on the relative structure of the 15 PTEs (subcomposition).

Based on the robust biplot analysis and taking into account the local geological setting and geochemical properties of elements, a sequential binary partition was performed using the 15 PTEs subcomposition. We were particularly interested in the main contamination sources and spatial patterns of the PTEs, which can be characterised using 4 specific balances (Z1, Z6-Z8). Each of these had elements belonging to distinct groups on the robust biplot, reflecting different geogenic and anthropogenic processes.

Interpolated maps were generated from these balances using the multifractal inverse distance weighted (MIDW) method (Cheng et al., 1994) with GeoDAS software. For the interpolation, we used a 5 km search radius and a minimum of 4 data points, with the map resolution set to 1 km based on the break point of the nearest distance cumulative plot of the sampling sites. Considering that the balance values can be negative and therefore not “log transformable”, a min-max normalisation was applied by scaling the original data within a specified range of values (e.g., ranging from 1 to 100). Min-max normalisation is a linear transformation applied to the original data that does not affect the geometrical structure (Han

and Kamber, 2001). After interpolation, the normalised values were back-transformed to visualize the original range of balances.

In addition, the concentration–area (C–A) fractal plot (Cheng et al., 1994) was used to classify the interpolated balance maps (Z1, Z6-Z8) and capture the different spatial patterns.. Profiles were also made for the interpolated balance maps in different directions to capture spatial variability and define outliers of specific groups of elements.

3.2.3. Compositional Abundance maps and the Soil Contamination Compositional Index

Soil contamination by PTEs represents one of the foremost environmental issues in the world today. High concentrations of PTEs and associated compounds put human beings at risk through dermal contact, inhalation and ingestion (Lim et al., 2008; Ji et al., 2008; Varrica et al., 2014). A compositional index should be invariant from a change in measurement units or multiplication by a constant (Egozcue, 2009). Compositional data can be properly analysed using log-contrasts as a balance between certain groups of elements, as is the case for pollution indices that consider the ratio of the geometric mean of pollutant concentrations with that of non-pollutant concentrations (Jarauta-Bragulat et al., 2015). Jarauta-Bragulat et al. (2015) also introduced a new air quality index (AQI) that was a good approximation of a log-contrast as a balance.

AQI was calculated as $AQI(x) = k * g_{n-1}(x)$, where k^* is a constant depending on the units of pollutant concentrations and $g_{n-1}(x)$ is the geometric mean of air pollutant concentrations. In order to get a normalised index, they used an arbitrarily large geometric mean value ($\max g_{n-1}(x)$) as a denominator which correspond to an almost incredible pollution of the air and modified the equation to be the following:

$$AQI^* = (k^* / \max g_{n-1}(x)) * g_{n-1}(x) \quad (1)$$

In this study, the same equation as Eq. 1 was tested on the 15 PTEs but we used the maximum geometric mean of the PTEs as a denominator and changed the name to Compositional Abundance Index (CAI). Compositional abundance indices were also calculated for specific subgroups that were categorized as the following: Volcanic (As, Be, Tl, V and Se), Terra Rossa soil and Siliciclastic (Cr, Ni, Co and Cd) and Anthropogenic (Pb, Sb, Sn and Zn) Compositional Abundance Indices. The calculation was applied to subgroups based on a sequential binary partition table using Eq. 1. We used 100 as the k^* constant in Eq. 1 to get abundance ranges between 0 and 100. Although the CAI is compositionally meaningful and easily interpretable, the maximum geometric mean of elemental concentrations in the denominator makes it a one-off estimator for a survey. This was the motivation for elaborating a new

Soil Contamination Compositional Index (SCCI) for both the total and subgroups of the 15 PTEs. The SCCI is stable and retains the same value for older observations even if the geochemical sampling is expanded. The SCCI was calculated using the following equation:

$$SCCI = g_{n-1}(x) / g_{n-1}(x_{bas}) \quad (2)$$

where the geometric mean of the 15 PTEs concentrations is $g_{n-1}(x)$ and the geometric mean of the baseline values of these elements is $g_{n-1}(x_{bas})$. We used the baseline values of each of the 15 PTEs published by De Vivo et al. (2016) for the entire Campania Region. These baseline values for each element were obtained by the Spectrum-Area (S-A) fractal method developed by Cheng et al. (1994, 2000). Although baseline values can be decomposed into anthropogenic and geogenic parts, the larger the SCCI is, the larger the contamination of topsoils may be. SCCI values close to 1 indicate little or no contamination by the investigated PTEs. The SCCI was also calculated for the same specific subgroups (Volcanic, Terra Rossa soil and Siliciclastic, Anthropogenic) using the respective baseline values to reveal contamination source patterns. The SCCI ratio values were interpolated using the MIDW method and classified by C-A fractal plots.

4. Results and discussion

4.1. Univariate statistical analysis

Table 1 displays the summary statistics of the 15 PTEs and the percentages below the detection limit for Cr, Sb and Se. In terms of variability, Cu has the highest robust coefficient of variation (CVR) (56.45%), followed by Se (50%) and Hg (46.15%). The lowest CVR is represented by Ni (26.45%), Co (26.67%) and As (26.89%). The distribution of variables is mainly right-skewed and not symmetric based on positive skewness and kurtosis values (Table 1).

Table 1. Summary statistics of the 15 investigated PTEs based on log-transformed data that was then back-transformed.

PTEs	Min	Q1	Mean	Mean-log	Median	Q3	Max	CV (%)	MAD	Skewness	Kurtosis	<DL (%)
Cu	2.51	31.89	97.30	63.45	58.57	124.19	2394.33	56.45	33.25	0.19	-0.16	
Pb	3.12	32.07	67.09	49.09	51.50	73.34	1305.91	39.16	20.23	0.12	1.32	
Zn	11.40	69.10	117.25	96.36	89.60	126.75	3210.60	28.76	25.70	0.80	2.77	
Ni	0.50	12.10	18.39	15.22	15.50	20.90	98.90	26.45	4.10	-0.48	1.35	
Co	0.80	7.60	10.81	9.83	10.50	13.30	79.00	26.67	2.80	-0.49	1.31	
As	0.60	8.10	11.97	10.51	11.90	14.80	111.50	26.89	3.20	-0.89	1.99	
V	5.00	43.00	63.71	57.52	59.00	84.00	224.00	32.20	19.00	-0.65	0.65	
Cr	0.325	11.10	20.20	15.73	16.00	23.10	808.40	35.00	5.60	-0.16	2.99	0.04
Cd	0.02	0.23	0.49	0.37	0.37	0.56	11.06	43.24	0.16	0.27	1.24	
Tl	0.05	0.74	1.37	1.06	1.40	1.95	3.62	42.86	0.60	-1.24	0.87	
Sb	0.013	0.43	0.97	0.65	0.65	0.98	40.79	40.00	0.27	0.14	3.43	0.22
Hg	0.003	0.033	0.094	0.058	0.052	0.098	6.775	46.15	0.025	0.50	0.97	
Se	0.065	0.20	0.43	0.33	0.40	0.60	2.40	50.00	0.20	-1.04	0.46	5.11
Sn	0.20	1.90	3.97	2.98	3.30	4.70	125.60	42.42	1.40	-0.27	0.63	
Be	0.20	2.70	4.43	3.71	4.50	5.90	16.90	33.33	1.50	-1.21	1.29	

The Tukey boxplots also reveal the asymmetry in the distribution of variables and the dominance of high outliers on a log-scale (Fig. 1). The medians of variables of the Campanian topsoil data exceed the European counterparts (GEMAS, Reimann et al., 2014) for all of the elements except for Cr. The elements with the highest ratios of medians are TI ($r=11.6$), Be and Sn ($r=4.58$) (Fig. 2), while the elements with the lowest ratios of medians are Cr ($r=0.8$) and Ni ($r=1.03$). All of the 15 PTEs have concentration values above the intervention limits for residential areas and a few samples also exceed the industrial limits set by Italian law (D. Lgs. 152/2006) (Fig. 1). The analyses for Cu, Zn, Pb, As, Be, Sn and TI are the most concerning because the vast majority of high outliers are above the residential intervention limits, indicating that there are potential risks for human beings (Fig. 1). It should be emphasised that Sn becomes toxic only if combined with organics to form organo-Sn compounds (Kabata-Pendias, 2011), though Italian Law reports it as a PTE. The average upper-crustal concentrations of the 15 PTEs (Wedepohl, 1995) are generally lower than the medians of the respective elements in the Campania Region with the exception of Cr, Ni and Co (Fig. 1).

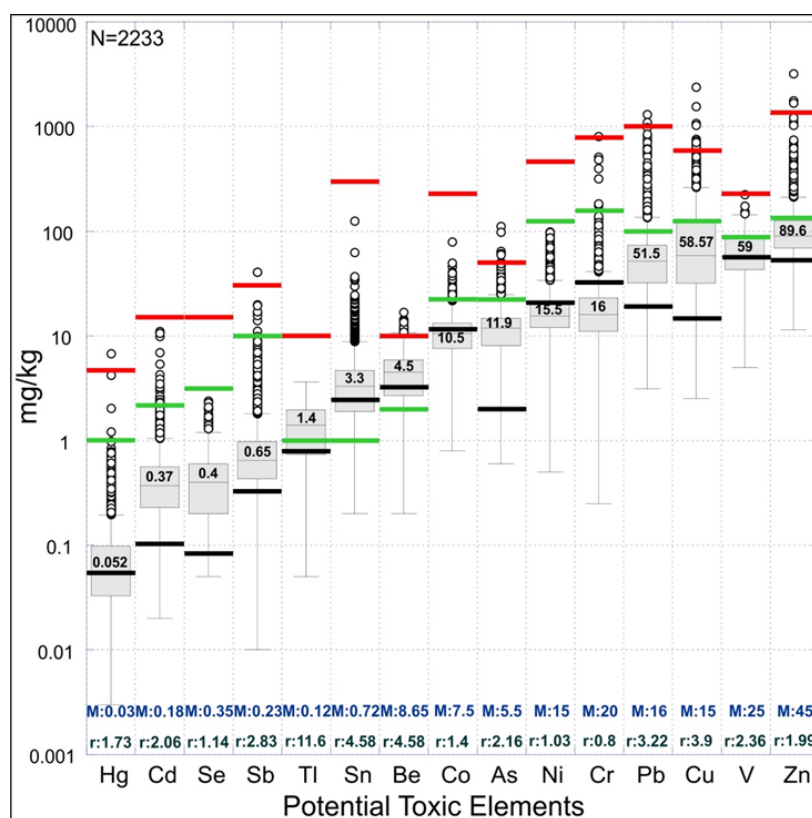
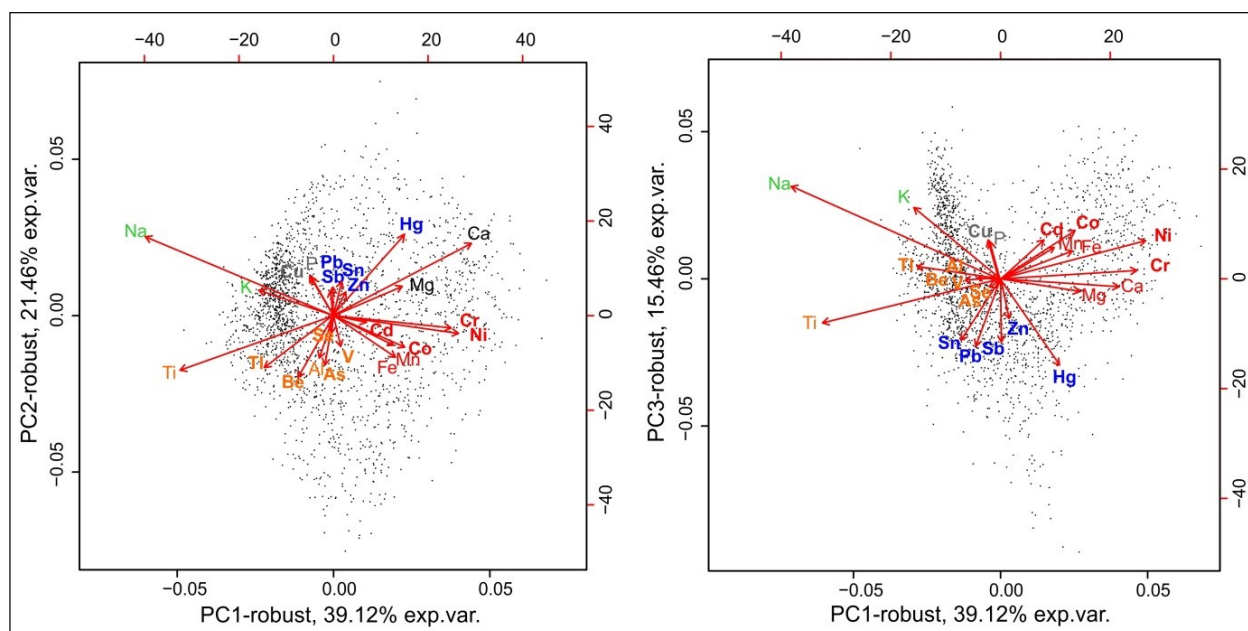


Fig. 1. Box-plots of the 15 PTEs. Green and red lines show the intervention limits for residential and industrial areas, respectively (D. Lgs. 152/2006). The thick black lines mark the average upper crustal concentration of the given element (Wedepohl, 1995). The numbers in the boxes are the medians of the respective elements. M: medians of respective elements from the Geochemical Mapping of European Agricultural Soil project (Reimann et al., 2014). r: ratios between the medians of the Campanian PTEs and the European counterparts.

4.2.1. Robust biplot analysis

Group A consists of the clr (centred log-ratio) coefficients of Hg, Pb, Sb and Sn, with rays pointing in the same direction with the highest loading of the clr-Hg (Fig. 2). This group is represented by clr coefficients that are likely related to anthropogenic activity including heavy traffic, fossil fuel combustion, illegal waste disposal and extensive agricultural activity. The anthropogenic origin of these elements was also revealed in stream sediment data by Albanese et al. (2007) and in other local-scale studies by Cicchella et al. (2005, 2016) and Petrik et al. (2018a). The clr-Hg might also be influenced by a combination of processes because the vertex deviates from the rest of Group A and reveals a high clr-variability (vector length) (Fig. 2). High Hg concentration has been attributed to anthropogenic (e.g., traffic, fossil fuel combustion) and hydrothermal activity (Cicchella et al., 2005; De Vivo et al., 2016).



Group B is represented by the clr-coefficients of 4 PTEs (Ni, Cr, Co and Cd) and 2 major elements (Fe and Mn) (Fig. 2A). The rays are pointing in the same direction with the longest vector length of Cr and Ni

(Fig. 2A). Cd has the lowest clr-variability of the 4 PTEs. The association of these 4 PTEs with Fe and Mn may be related to phyllosilicates and Fe-/Mn-sulphides from black shaley facies of siliciclastic rocks (e.g., Crete Nere Fm.). However, the biplot projected on the PC1-PC3 axes reveals that these 4 PTEs are also accompanied by the major elements of Ca and Mg (Fig. 2B). This association reflects different geogenic processes that may be indicative of elements sourced from Terra Rossa soils. Terra Rossa soils (e.g., Haplic Luvisols) were mapped (Di Gennaro et al., 2002) throughout the NW-SE trending carbonate massifs of the southern Apennines in the Campania Region where some bauxite mineralisation was also found (Mondillo et al., 2011; Boni et al., 2013).

In fact, the 4 PTEs of Group B seem to be associated not only with shaley facies of the siliciclastic rocks but also with the Terra Rossa soils. Group C forms a separate association that includes the clr-coefficients of As, Be, Ti, V and Se and the major elements of Al and Ti. The rays are scattered but all point in the same direction with the longest vector length of Ti (Fig. 2A). These clr-coefficients may be related to the presence of volcanoclastic rocks and hydrothermal/volcanic activity (Tarzia et al., 2002; Lima et al., 2003b; De Vivo et al., 2016). However, the relative abundance and spatial pattern analysis of As and Be revealed that these elements may be at least partially associated with advanced pedogenic processes (e.g., mature Andisols found with a Chemical Index of Alteration of 70–80) in the Campania Region (Petrik et al., 2018b; c). De Vivo et al. (2016) related some of these elements (e.g., Al, As, Be and Ti) to the oldest volcanic products of the Campanian eruption events.

Group D contains only the clr of Cu and the major element of P, with vertices separated from other associations that point in a different direction. Studies of the stream sediment and topsoil from the Campania Region and Sarno River Basin demonstrated that these elements are related to agricultural activity using sewage sludge and phosphate fertilizers (Albanese et al., 2007; Adamo et al., 2014; Cicchella et al., 2016).

Finally, the last association includes the clr-coefficients of Na and K that are not reported as PTEs by Italian law (D. Lgs. 152/2006) and have high abundances related to the youngest volcanic rocks of Mt. Somma-Vesuvius and the Phlegraean Fields (De Vivo et al., 2016).

4.2.2. Balance maps (Z1, Z6-Z8)

Based on the robust biplot analysis, a sequential binary partition was performed that involved the 15 PTEs subcomposition (Table 2). There was a particular interest in groups of elements that represented different geogenic and/or anthropogenic processes. For this reason, 4 balance maps (Z1, Z6-Z8) were further analysed to reveal spatial patterns and contamination sources (Table 2).

Table 2. Sequential binary partition table of the 15 PTEs subcomposition to obtain balances (Z1-Z14). Positive and negative signs (parts) in the i-th order partition show the separated two parts (groups of elements) used in the calculation of the given balance. The empty entries are parts (elements) that were not involved in the partition in the i-th order. Balances were calculated with CoDA-Pack 2.02.21. The highlighted balances were further analysed for the identification of spatial patterns.

Elements	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14
As	+	+	+	+										
Be	+	+	+	-										
Tl	+	+	-											
V	+	-			+									
Se	+	-			-									
Pb	-					+	-	-	-	-	-			
Sb	-					+	-	-	-	-	+			
Sn	-					+	-	-	-	+				
Zn	-					+	-	-	+					
Hg	-					+	-	+						
Cu	-					+	+							
Cr	-					-						+	+	
Ni	-					-						+	-	
Co	-					-						-		+
Cd	-					-						-		-

4.2.2.1. Z1 (ilr-1) balance map

The Z1 balance map illustrates that the abundance of the volcanic rock elements (Group C: As, Be, Tl, V and Se) with respect to all of the other PTEs (Fig. 3A, Table 2). The vast majority of Z1 balance values are below 0, meaning that the dominance of volcanic rock elements is lower than other PTEs in the study area (Fig. 3A). The highest spatial abundance of volcanic rock elements (-0.04–0.88) is related to large volcanic complexes (e.g., Roccamonfina, Phlegraean Fields and Mt. Somma-Vesuvius) and carbonate massifs (e.g., Mt. Picentini and Mt. Cervati) overlain by pyroclastic deposits (Figs 3A, C-Profiles 1 and 2).

The lowest spatial abundance (-3.28 to -1.87) was from topsoils found over siliciclastic deposits in the north-eastern and southwestern parts of the Campania Region (Fig. 3A). Low spatial abundances of volcanic rock elements were also revealed along large rivers like the Volturno, Sele and Calore, where other non-volcanic elements have accumulated by river transport thereby disturbing the spatial distribution pattern (Figs 3A, C-Profiles 1 and 2). Interestingly, large urbanised areas like Naples and Salerno have a low abundance of volcanic rock elements despite being located close to large volcanic centres (Fig. 3A, C-Profiles 1 and 2). The reason for this unexpected finding is that these city soils have different contamination sources (e.g., heavy traffic, industrial centres) that might affect the spatial pattern. Profile 2 reveals a SW to NE decrease in volcanic rock elements extending from the volcanic centres towards the siliciclastic zone of the southern Apennines (Fig. 3C).

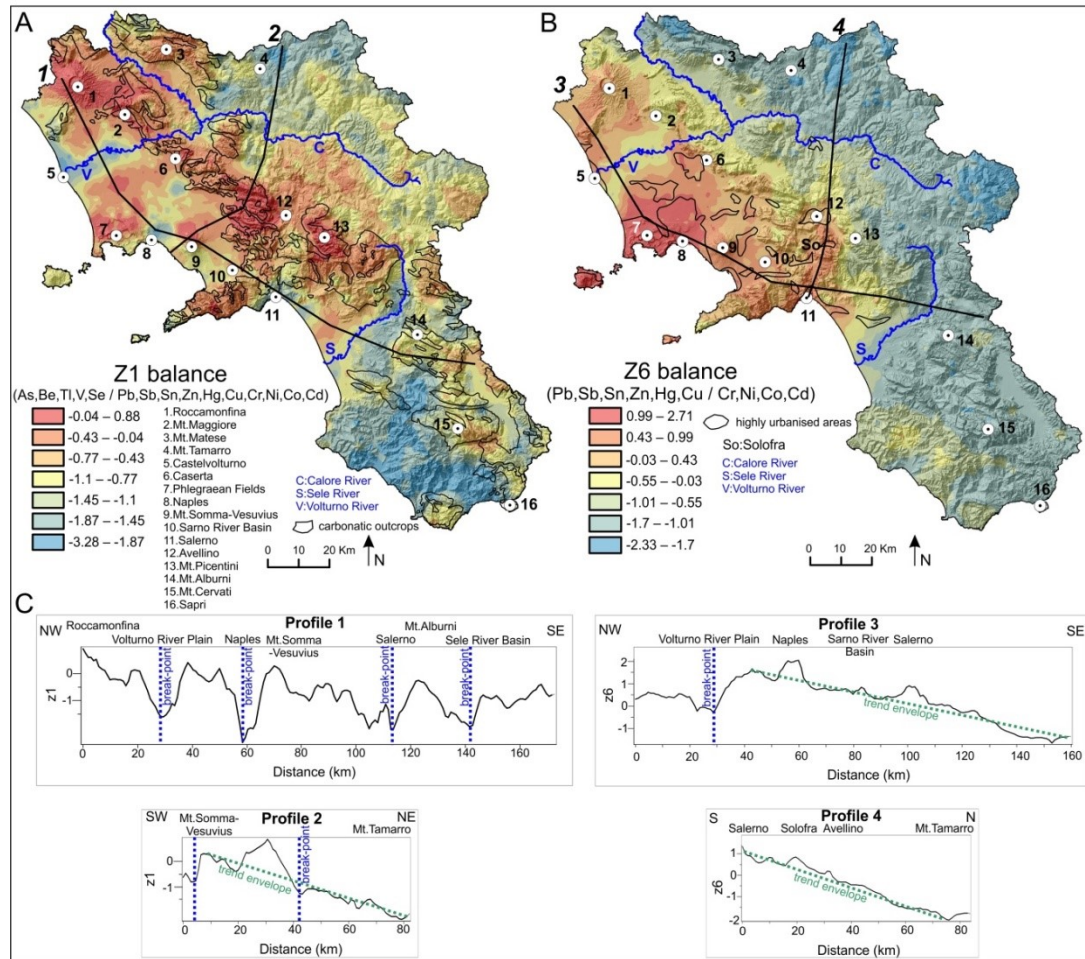


Fig. 3. (A) Z1 balance map showing the spatial abundance of volcanic rock elements (Group C: As, Be, Ti, V and Se) compared to other PTEs. (B) Z6 balance map demonstrating the spatial abundance of Terra Rossa soil and Siliciclastic rock elements (Group B: Cr, Ni, Co and Cd) compared to Group A (Pb, Sb, Sn, Zn and Hg) and Group D (Cu). (C) Profiles made in directions relevant for capturing the spatial variability and trend of abundances in different subgroups of the elements. The profile numbers and lines are shown on the maps (A and B). The sequential binary partition of the given balances can be seen in Table 2.

4.2.2.2. (ilr-6) balance map

The Z6 balance map represents the spatial abundance of Terra Rossa soil and Siliciclastic rock elements (Group B: Cr, Ni, Co and Cd) compared to other PTEs that primarily have anthropogenic contamination sources (Groups A and D) (Fig. 3B, Table 2). The range of balance values are quite symmetric (-2.33–2.71) indicating that no element group is dominant between Group B and Groups A and D. The negative balance values indicate that the spatial abundance of Terra Rossa soil and Siliciclastic rock elements (Group B) exceeds that of elements belonging to Groups A and D (Fig. 3B). The highest spatial abundance of Group B (-2.33–-1.7) can be observed in the eastern part of the Campania Region where siliciclastic rocks including shaley facies of the Crete Nere Fm. are dominant. The presence of Co, Cr and Ni was attributed to siliciclastic rocks and alteration from weathering and co-precipitation processes by Albanese et al. (2007) and Buccianti et al. (2015). The other high spatial

dominance of Group B can be observed overlapping some of the large carbonate massifs where Terra Rossa soils (e.g., Haplic Luvisols) with bauxite mineralisation (Mondillo et al., 2011) are present (Fig. 3B). High concentrations of Co, Cr, Ni and Cd have also been observed in Terra Rossa soils located in Slovenia and Croatia (Durn et al., 1999; 2001; Miko et al., 1999). There is a distinct separation between high positive balance values associated with elements that mainly have an anthropogenic origin in the Campania Plain area (Groups A and D) and the rest of the region where urbanisation is less dominant (Fig. 3B). Profiles 3 and 4 clearly show the decrease in balances extending away from the highly urbanised Campania Plain area towards the southern Apennines (Fig. 3C).

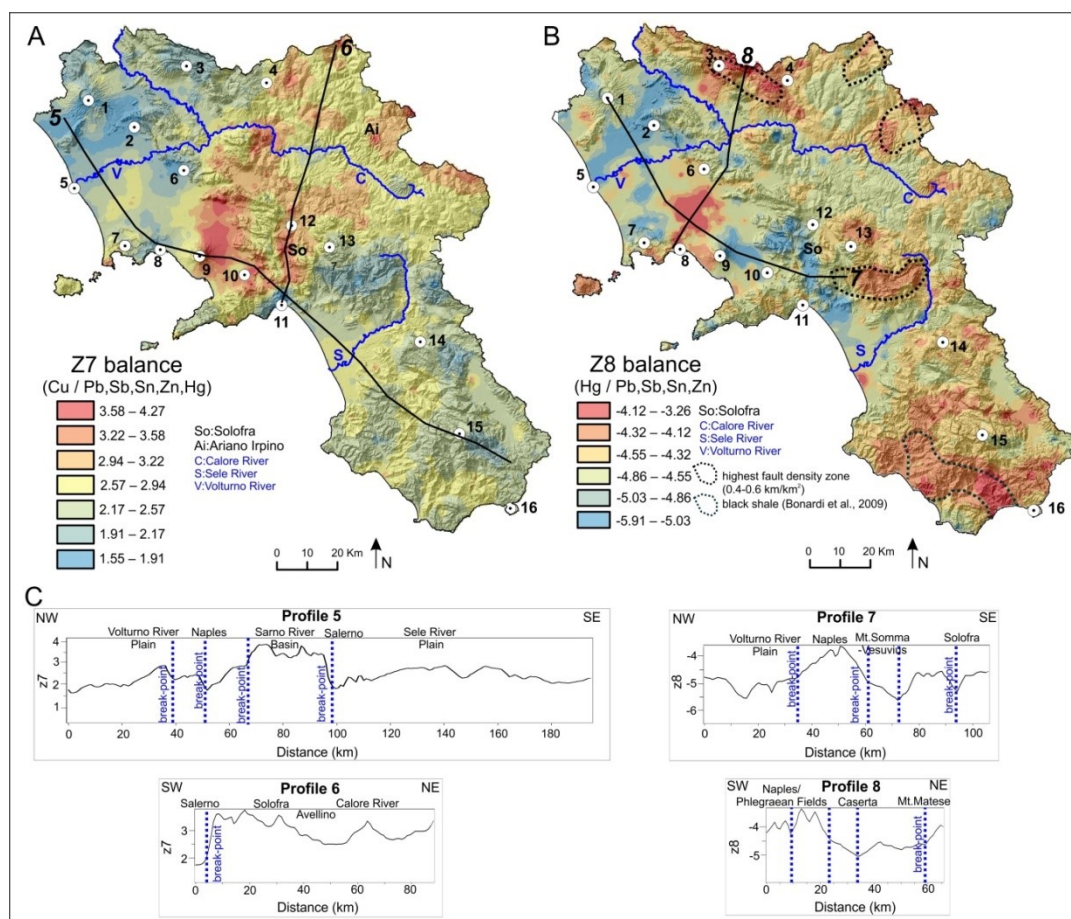


Fig. 4. (A) Z7 balance map showing the spatial abundance of Cu (Group D) compared to elements of Group A (Pb, Sb, Sn, Zn and Hg). (B) Z8 balance map demonstrating the spatial abundance of Hg with respect to other elements belonging to Group A (Pb, Sb, Sn and Zn). (C) Profiles made in directions relevant for capturing the spatial variability and trend of abundances in different subgroups of the elements. The profile numbers and lines are shown on the maps (A and B). The sequential binary partition of the given balances can be seen in Table 2. The topographic names indicated by 1–16 are the same as in Fig. 3A.

4.2.2.3. Z7 (ilr-7) balance map

The Z7 balance map reveals the spatial dominance of Cu compared to other PTEs that belong to Group A (Fig. 4A, Table 2). The positive balance values show the abundance of Cu with respect to the other elements. The highest spatial abundance of Cu is related to the slopes of Mt. Somma-Vesuvius and the Sarno River Basin where expansive agricultural areas utilize fertilisers and sewage sludge containing Cu compounds. High Cu concentrations are observed in scattered patches in the eastern part of the Campania Region around Ariano Irpino where there are numerous olive orchards and vineyards. Cicchella et al. (2016) also emphasised that intensive agriculture in the Sarno River Basin is a primary contamination source for Cu and P.

Profile 5 clearly shows large peaks (higher spatial abundance of Cu) in the Sarno River Basin and reveals two large depressions in highly urbanised regions (Naples and Salerno) where other anthropogenic elements of Group A are abundant (Fig. 4C). Profile 6 unveils a large break in the data northeast of Salerno and a progressive decrease in Z7 balance values toward Avellino, suggesting that elements of other anthropogenic origins (Group A) have become more dominant (Fig. 4C).

4.2.2.4. Z8 (ilr-8) balance map

The Z8 balance map compares the spatial abundance of Hg with other elements of Group A (Fig. 4B, Table 2). The negative balance values shown in Fig. 4B indicate that other elements of Group A (Pb, Sb, Sn and Zn) have a higher abundance than Hg. The highest dominance of Hg (-3.26 to -4.12) is not spatially coherent with other data suggesting that there are multiple processes governing the concentration of Hg, which was also confirmed by the robust biplot analysis (Fig. 2, Tables 1 and 2). The high spatial abundance of Hg is associated with the most urbanised areas between Naples and Caserta (Fig. 4B), where there is intensive vehicle traffic and oil refineries are present. The high Hg dominance can be seen in the Phlegraean Fields and the island of Ischia, and is attributed to hydrothermal activity of volcanic centres (Tarzia et al., 2002; Lima et al., 2003b; De Vivo et al., 2016). High Hg abundances are also observed in the southwestern part of the Campania Region, where there are outcrops of black shales known to contain higher concentrations of Hg (~0.18 mg/kg; Reimann and de Caritat, 1998). Other high abundances of Hg might be sourced from dense fault networks (0.4-0.6 km/km²) and related hydrothermal activities close to the Mt. Matese, where the highest precipitation rates also occur (1500-2000 mm/y). High precipitation rates and the formation of organic-rich topsoils were considered to be critical factors for the accumulation of Hg in Norway (Ottesen et al., 2013).

High abundances of other elements of Group A (Pb, Sb, Sn and Zn) are represented by large negative balance values (-5.03—-5.91) occurring between Salerno and Avellino, where the large Solofra industrial district is located (Fig. 4B). High negative balances can also be observed around the large volcanic complexes of Roccamonfina and Mt. Somma-Vesuvius where, besides some local anthropogenic

sources, geogenic processes (e.g., magmatic differentiation) control the enrichment of incompatible elements and PTEs in topsoil formed from volcanic rocks (Giannetti and Luhr, 1983; Ayuso et al., 1998; Peccerillo, 2005).

4.3. Compositional Abundance maps and SCCI

In order to assess the level of contamination in the Campania Region, contamination abundance indices were calculated for the 15 PTEs and subgroups using Eq.1.

The Compositional Abundance Index (CAI) can be interpreted as the relative difference in the spatial dominance of the PTEs, so it reflects the different amounts of enrichment or depletion in the study area. In addition, the calculation of the new SCCI (Soil Contamination Compositional Index) using Eq. 3 made it possible to quantify the contamination level of PTEs and reveal the contamination source patterns of different subgroups.

4.3.1. Total Compositional Abundance Index map

The Total Compositional Abundance Index map shows that the highest spatial abundance of PTEs (54.43-82.4) occurs in the highly urbanised areas of Naples and Salerno, where there is the highest concentration of heavy traffic and industrial activity (Fig. 5A). The high spatial abundance of PTEs is also associated with the Sarno River Basin, which is one of the highest contaminated river basins in Italy due to a high level of industrial and agricultural activity (Cicchella et al., 2016) (Fig. 5A). Adamo et al. (2014) emphasised that even if the concentrations of some of the PTEs exceed the intervention limits, the bioavailability of these elements is low meaning that there are minimal risks for humans in the Sarno River Basin.

The industrial district of Solofra between Salerno and Avellino also has a high spatial abundance of PTEs (54.43-82.4) related to the tannery industry, which represents one of the main contamination sources in the area (Fig. 5A). The highest SCCI values (1.25-1.92) suggest that the above mentioned areas are highly contaminated (Fig. 5A). Indeed, the SCCI values indicate that the geometric means of the 15 PTEs can be almost twice as high as those of the respective baseline values in these areas (SCCI: 1.51-1.92, Fig. 6A). The heavy urbanization, high industrial activity and dense road networks that define this part of the Campania Region are the sources for these results.

In contrast, other high spatial abundances of PTEs (54.43-82.4) in the Mt. Matese and the area between Sapri and Mt. Alburni are mainly from geogenic sources (e.g., Terra Rossa soils) and pyroclastic material (Fig. 5A). However, low SCCI values (1.001-1.12) reveal no contamination in these areas except Sapri where the values are medium-high (1.12-1.51) (Fig. 6A). Thiombane et al. (2018a) suggested that the anthropogenic sources for Pb and Sb around urban areas like Sapri are frequent traffic jams. A low spatial dominance of PTEs (< 21.39) was identified in the eastern and southwestern

parts of the Campania Region where urbanisation and industrial activity is rather low (Fig. 5A). Some medium-high SCCI values (1.12-1.51) were also revealed in the middle of the low abundance zone of PTEs around Ariano Irpino, where moderate contamination of the soils is possibly related to agricultural activity.

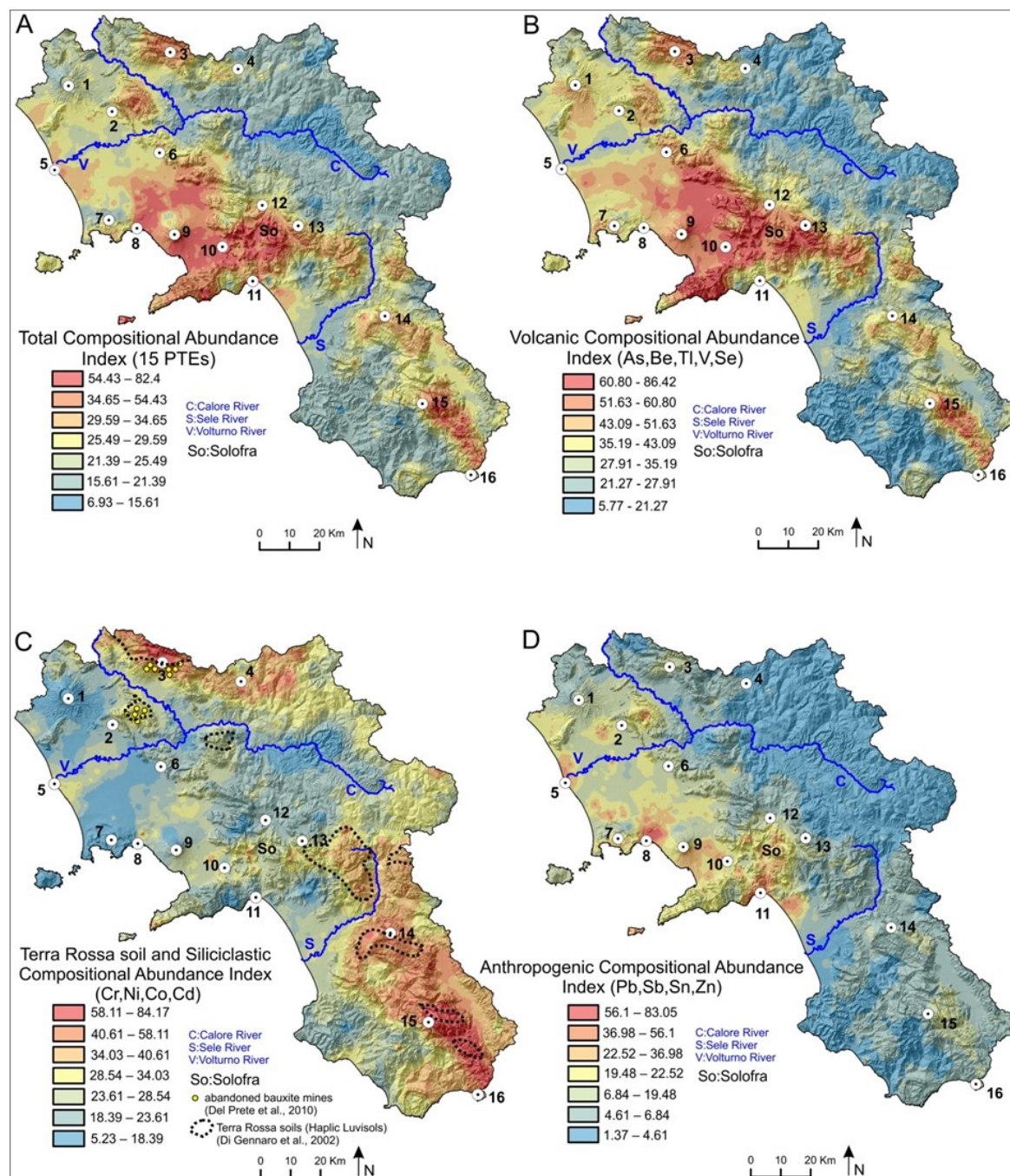


Fig. 5. (A) Total Compositional Abundance Index map showing the spatial abundance of the 15 PTEs. (B) Volcanic Compositional Abundance Index map illustrating the spatial dominance of volcanic elements (Group C). (C) Terra Rossa soil and Siliciclastic Compositional Abundance Index map demonstrating the spatial dominance of Group B elements (Cr, Ni, Co and Cd). (D) Anthropogenic Compositional Abundance Index map showing the spatial dominance of elements with anthropogenic sources (Group A).

4.3.2. Volcanic Compositional Abundance Index Map

The Volcanic Compositional Abundance Index map for Group C (As, Be, Se, Tl and V) shows the spatial abundance of volcanic rock elements (Fig. 5B). The highest dominance (60.8-86.42) is related to large volcanic centres, especially in the eastern and southern parts of Mt. Somma-Vesuvius (Fig. 5B).

Petrik et al. (2018b and c) revealed that there were high concentrations of Be and As in topsoils over pyroclastic rocks related to the Ottaviano and Avellino eruptions (Ayuso et al., 1998). The spatial extent of these two eruptions corresponds to the high spatial abundance of volcanic elements. The Volcanic Compositional Abundance Index map also reveals that the carbonate massifs overlain by pyroclastic deposits have a high spatial abundance (> 60.8) of volcanic elements (Fig. 5B).

The SCCI values of Group C are very low and close to 1 in most of the study area indicating that there is little to no contamination by these elements (Fig. 6B). The lowest SCCI values (1.0007-1.006) found around large metropolitan areas (e.g., Naples, Salerno, and Avellino) confirm that there are no anthropogenic contamination sources for these elements. Slightly elevated SCCI values (1.07-1.13) from Mt. Somma-Vesuvius and the Phlegraean Fields can be explained by the presence of various volcanic soils types over different pyroclastic levels.

The spatial coincidence of the highest abundance zones on the Volcanic and Total Compositional Abundance Index maps confirms that volcanic elements (Group A) certainly contribute to the high dominance of the 15 PTEs, especially around large volcanic centres (Figs. 5A and B). According to Albanese (2008), the bioavailability of these elements is less than 10% for urban soils. The lowest abundance of volcanic rock elements (< 27.91) is in the north-eastern and southwestern parts of the Campania Region where siliciclastic rocks are found without pyroclastic deposits (Fig. 5B).

4.3.3. Terra Rossa soil and Siliciclastic Compositional Abundance Index Map

This map reveals the spatial pattern and abundance of Terra Rossa soils and Siliciclastic rock elements (Group B: Co, Cr, Ni and Cd). The highest spatial abundance (58.11-84.17) is related to the carbonate massifs of the southern Apennines (e.g., Mt. Matese and Mt. Alburni) and in the area from Sapri to Mt. Cervati where Terra Rossa soils (e.g., Haplic Luvisols) with evidence for bauxite mineralisation can be found (Fig. 5C). The high dominance can also be observed east of the Calore River where siliciclastic rocks with shaley facies are present (e.g., Crete Nere Fm.) (Fig. 5C). However, these areas are characterised by very low SCCI values (< 1.01) that typically signify there are no sources of contamination (Fig. 7C). Cicchella et al. (2005) and Albanese et al. (2007) suggested that lithology (e.g., siliciclastic rocks) can play an important role in controlling the accumulation of these elements. However, high spatial abundance zones (> 58.11) were also found around large urbanised areas (e.g., Naples), industrial zones (e.g., Solofra) and at the mouths of the Sele and Volturno Rivers (Fig. 5C).

The highest SCCI values (1.29-1.82) reveal that these areas are highly contaminated by elements of Group B (Fig. 6C). These areas are characterised by high industrial activity (e.g., tannery industry in Solofra Basin: Cr and Ni), intensive agricultural activity (e.g., Sarno River Basin: Cd, Co and Cr) and heavy vehicle traffic (Cd). Extensive cultivation occurs along alluvial plains that border large rivers, but in the case of the Volturno River Plain, the role of co-precipitation induced by Fe- and Mn-oxy-hydroxides in soils of alluvial origin cannot be ruled out (Ducci et al., 2017).

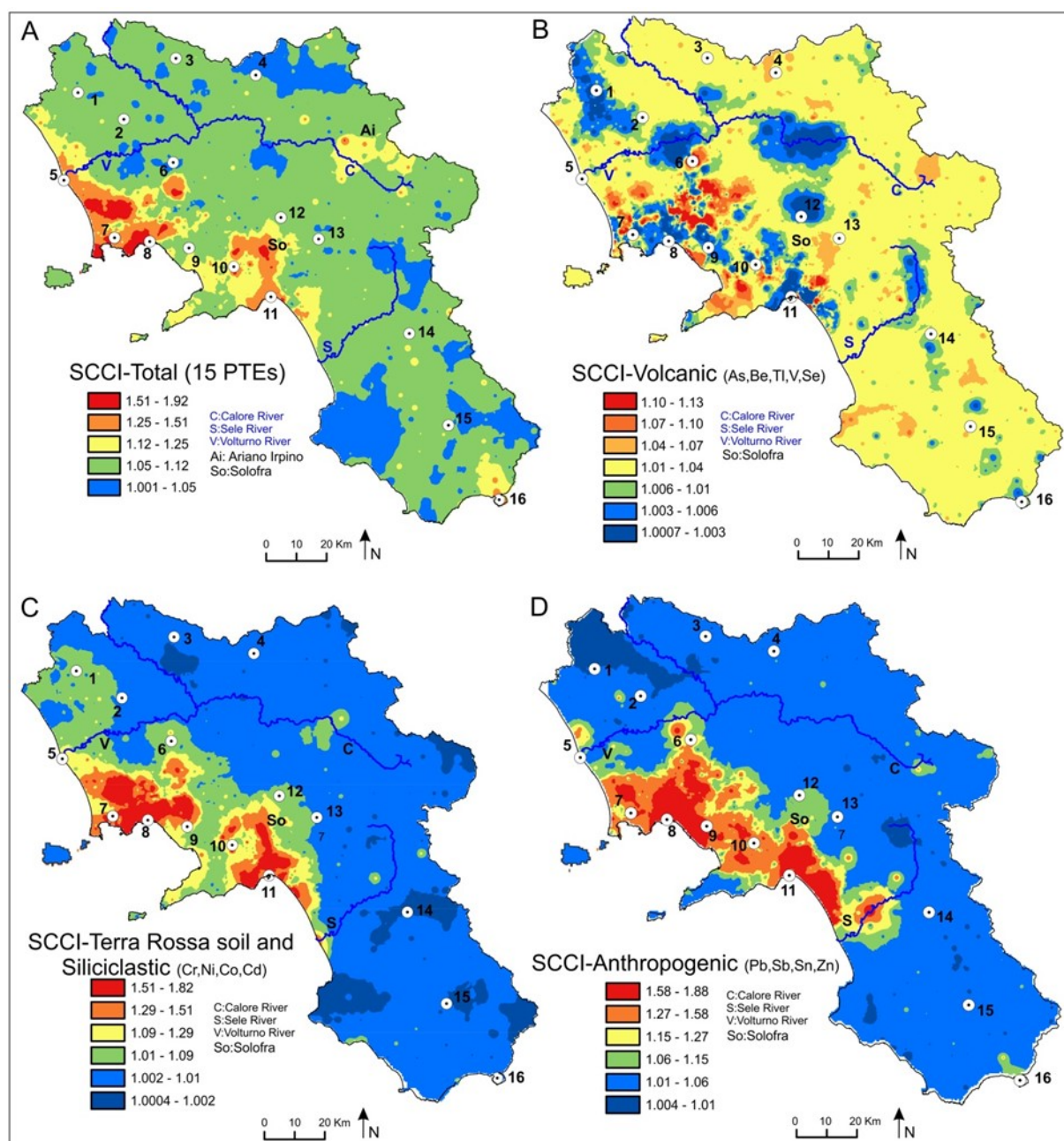


Fig. 6. (A) Soil Contamination Compositional Index (SCCI) map of the 15 PTEs. (B) The SCCI map of the volcanic rock elements (As, Be, Se, Ti and V). (C) The SCCI map of the Terra Rossa soil and Siliciclastic elements (Cr, Ni, Co and Cd). (D) The SCCI map of the anthropogenic elements (Pb, Sb, Sn and Zn). The SCCI was calculated using Eq. (2). The topographic names indicated by 1–16 are the same as in Fig. 3A.

4.3.4. Anthropogenic Compositional Abundance Index Map

Finally, we analysed the spatial abundance of elements that clearly have an anthropogenic origin (Group A: Pb, Sb, Sn and Zn). The Anthropogenic Compositional Abundance Index map unveils the highest spatial dominance (56.1-83.05) around highly urbanised areas like Naples and Salerno and in areas with extensive agricultural and industrial activity such as the Sarno River Basin and Solofra industrial district, respectively (Fig. 5D). The highest levels of contamination (SCCI: 1.58-1.88) occur around large metropolitan areas where the geometric means of the 15 PTEs can be twice as high as the counterpart baseline values (Fig. 6D). The primary role of the anthropogenic sources of Pb and Zn in this area was also described by Petrik et al (2018a). Some sites around Sapri have moderately high SCCI values (1.15-1.27), which likely reflect contributions from heavy traffic in the urbanised centre (Fig. 6D).

However, the geogenic contribution of pyroclastics from volcanic centres should also be considered as a source since elements of Group A (Pb, Sb, Sn and Zn) are also enriched from volcanic processes (Peccerillo, 2005).

The low dominance of the Anthropogenic Compositional Abundance Index (<19.48) was found in the eastern and southwestern parts of the Campania Region away from highly urbanised and industrial areas (Fig. 6D). These areas have very low SCCI values associated with Group A (<1.06) and therefore indicate that there is no anthropogenic contamination (Fig. 6D).

5. Conclusions

This study focuses on providing a better understanding of the compositional behaviour and relative structure of the 15 PTEs in the Campania Region, which is considered to be a highly contaminated region in Italy. Special emphasis was put on the relative abundance and ratios of specific subcompositions (balances) of PTEs by mapping spatial patterns and making profiles to reveal spatial variability.

Preliminary steps are provided for analysing the spatial abundance of PTEs by applying a compositional abundance index. In addition, a new soil contamination compositional index was elaborated to quantify the contamination of topsoil by the 15 PTEs and subgroups.

The high spatial abundance of the 15 PTEs is related to highly urbanised (Naples and Salerno), highly industrial (Solofra) and intensive agricultural areas (Sarno River Basin), where the anthropogenic subgroup (Pb, Sb, Sn and Zn) has a high dominance and the high SCCI values indicate that contamination is from anthropogenic sources. The high spatial dominance of the volcanic rock subgroup (As, Be, Se, Tl and V) in the same area is certainly related to geogenic factors associated with the presence of pyroclastic deposits. Although the high spatial abundances of Group B elements (Cd, Cr,

Co and Ni) are related to Terra Rossa soils and shaley facies of siliciclastic rocks of the southern Apennines, these elements can also reach high abundances and reflect contamination (i.e. high SCCI values) in urbanised (e.g., traffic) and industrialised areas (e.g., tanneries, alloy manufacturing). Other high spatial abundances of the 15 PTEs with little or no contamination (i.e. very low SCCI values) are associated with carbonate massifs that reflect a mixture of geogenic factors (weathering, advanced pedogenic processes, adsorption and co-precipitation with Fe-/Mn-oxy-hydroxides, presence of pyroclastic deposits) that might all be responsible for PTE accumulation.

The lowest spatial dominance of the 15 PTEs occurs in the north-eastern and southwestern Siliciclastic zones of the Campania Region where urbanisation and industrial activity is minimal and therefore contamination by any source can be excluded.

A future goal is to improve the Compositional Abundance Index and the Soil Contamination Compositional Index by giving weights to the PTEs based on the bioavailability and toxicity levels of each element.

Section 2.3

Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs)
through a compositional discrimination analysis: A case study on
the Campanian topsoil data

This section has been published in:
Journal of Geoderma
Volume 331, 1 December 2018, Pages 87-99

Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) through a compositional discrimination analysis: A case study on the Campanian topsoil data

Abstract

One of the main objectives of environmental geochemistry is to reveal the source and spatial patterns of different inorganic and organic elements/compounds with special emphasize on potentially toxic elements. In the last couple of decades, environmental geochemists mainly focused on the identification and separation of geochemical anomalies from background/baseline values of potentially toxic heavy metals and the estimation of their ecological and human health risks. However, the main concern with previously published papers on this issue is that the majority of them simply neglected the compositional nature of geochemical data; hence results became spurious and biased and can be interpreted with reservations. Our goal is to identify, interpret and discriminate the source patterns of 4 potentially toxic elements in the Campania Region (Italy) emphasising on their ratios and spatial abundance using multivariate compositional data analysis. This study contributes to understand the compositional behaviour and relative proportion of 4 potentially toxic elements whose precise contamination sources have been unclear before. A workflow of compositional data analysis including a new discrimination index has been elaborated to identify their possible sources of contamination and enrichment. The investigated data set includes 18 elements derived from 3669 topsoil samples, collected at an average sampling density of 1 site per 3.2 km². First, robust biplots and factor analysis were performed to get an overview of elemental associations and reduce the dimensionality of the data set. They revealed that the 4 PTEs belong to different groups. The multivariate regression analysis using alr-transformed (additive logratio) data proved the strong linear relationship between Fe, Mn (independent) and each investigated PTE (dependent), but also unveiled deviation trends in case of Zn and Pb. Based on the multivariate regression result, a sequential binary partition was performed by means of 6 variables (the 4 PTEs, Fe, and Mn) to obtain balances. Balances are ilr-coordinates (isometric-logratio) which can be interpreted as ratios of specific groups of elements. They were used to generate interpolated maps by using multifractal method to see spatial patterns and proportions of elemental associations. A new index has been elaborated based on the bivariate regression of balances and their standardised residuals, which was particularly useful to identify and separate the sources of anthropogenic contamination and geogenic enrichment of respective elemental associations. The large urban and industrial areas (e.g. Naples, Salerno) along the coastline are mainly contaminated by Pb and Zn due to heavy traffic and alloy production. Some Cr and Ni contamination was discerned in the Sarno Basin where the Solofra industrial district is likely to be the principal source through releases from tannery industry. The large

volcanic complexes (e.g. Mt. Somma-Vesuvius, Phlegraean Fields, Mt. Roccamonfina) are all characterised by geogenic enrichment of Zn and Pb. In contrast, Cr and Ni-geogenic enrichment is mainly related to the siliciclastic deposits.

1. Introduction

Geochemical surveys aim to enhance mineralisation and contamination through exploration and environmental geochemistry, respectively. For that, various geostatistical computations have been used to identify source patterns of different elements related to underlying geological features and/or anthropogenic activities (Cheng et al., 1994; Cheng, 1999; Lima et al., 2003a; Reimann and De Caritat, 2005a). Graphical and statistical methods (e.g. classification using boxplots or cumulative probability plots) have been elaborated to study the elemental distribution and identify univariate outliers (Tennant and White, 1959; Sinclair, 1976, 1983; Kürzl, 1988; Reimann et al., 2008).

The detection of multivariate outliers using certain cut-off values on the Mahalanobis distance and observed covariance ratio were also successfully applied on compositional data (Bollen, 1987; Roosseeuw and Van Zomeren, 1990; Filzmoser et al., 2012; Buccianti et al., 2015).

Background/anomaly separation by using various fractal methods (e.g. Concentration-Area) was given particular attention in mineral exploration, environmental health-risk analysis and regional topsoil studies (Cheng et al., 1994; Cheng, 1999; Agterberg, 2001; Lima et al., 2003a; Fabian et al., 2014; Albanese et al., 2015; Zuo et al., 2015). In addition, the question of anthropogenic or geogenic origin has always been in the interest of investigations. Several indices have been invented to separate anthropogenic contamination from geogenic background values like Enrichment Factor (EF, Chester and Stoner, 1973) or Geoaccumulation Index (Müller, 1979). These indices were critically reviewed by Reimann and de Caritat (2005) claiming that their values vary and depend on the different parent rock materials and chosen reference media as well as reference elements. In addition, these indices do not take into account the different biogeochemical processes which may have remarkable impact on elemental enrichment/contamination (Reimann and de Caritat, 2005). Finally, geochemical data are inherently compositional data and indices based on the absolute values of elements disregard the compositional principles like scaling invariance and subcompositional coherence (Pawlowsky-Glahn and Buccianti, 2011).

In this study, the main objective is to identify, interpret and map the source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) on a high-density Campanian topsoil data using multivariate compositional data analysis. A new compositional discrimination index has also been elaborated taking into account the compositional nature of geochemical data to better understand and distinguish the precise contamination/ enrichment sources of the 4 PTEs which have been unclear before.

2. Methods

A workflow of compositional data analysis has been elaborated to obtain additional information to discriminate the sources of the 4 PTEs in the study area (Fig. 1).

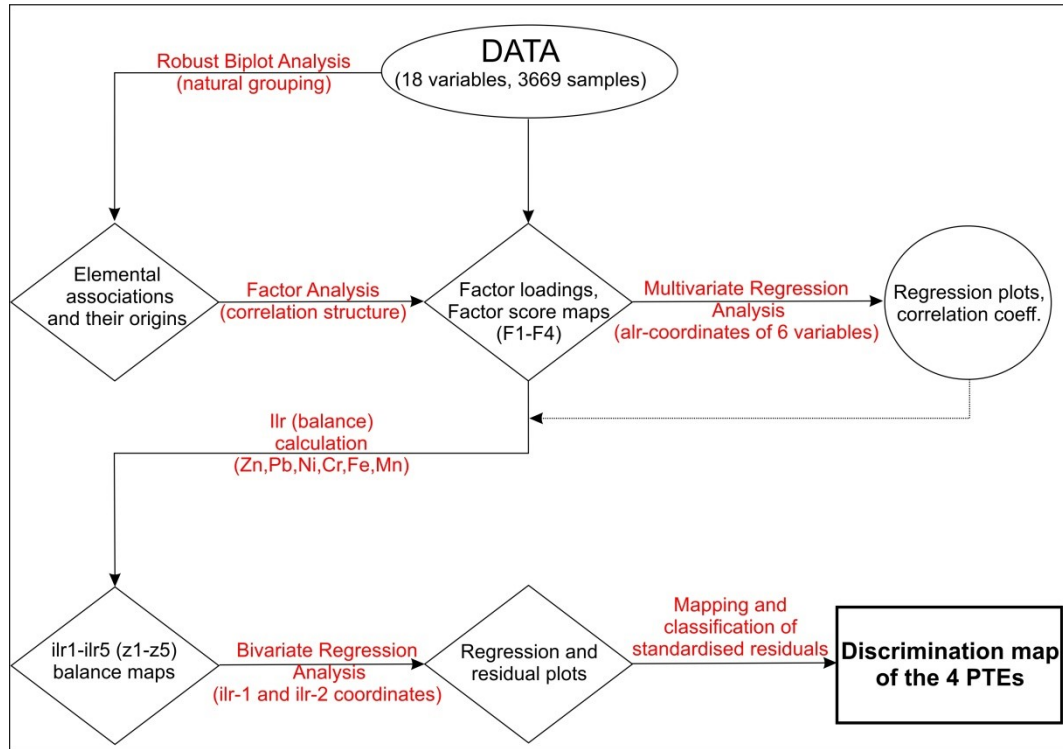


Fig. 1. The flowchart of the compositional discrimination analysis.

To better visualize the element distributions and quantify possible natural or anthropogenic impact, robust compositional biplots were generated. The section 2.1 fully describes compositional biplots and their practical usage.

The number of all measured elements (53) was reduced to 18 (Table 1) based on 3 main criteria (see section 2.1). Based on such criteria, in this paper, we present biplots with the reduced number of variables (18, Table 1) used as inputs for factor analysis.

Factor analysis (FA) was the multivariate statistical tool to explain the correlation structure of the variables through few numbers of factors (Reimann et al., 2002). Furthermore, FA can be successfully used to reveal the elements sources related to their main hypothetical origins.

To minimize or eliminate the presence of outliers and spurious correlation (Pawlowsky-Glahn and Buccianti, 2011), isometric logratio transformation (ilr) was applied on raw-data prior to multivariate analysis (Filzmoser et al., 2009b).

Table 1. Descriptive statistics of the 18 elements (mg/kg) using 3669 topsoil samples from the Campania Region, Italy. Q1 and Q3 indicate the 1st and 3rd quartiles, respectively.

Variables	Minimum	Q1	Mean	Median	Q3	Maximum	Std. Deviation	Skewness	Kurtosis
Cu	2.51	32.28	109.32	62.24	128.25	2394.33	156.83	5.49	44.56
Pb	3.12	36.86	73.85	54.17	78.52	2052.18	91.72	7.71	100.16
Zn	11.4	69.70	119.30	91	127.95	3210.60	118.46	9.23	164.68
Ni	0.5	10.00	16.23	14.80	18.30	100.90	10.98	2.42	9.97
Co	0.5	6.90	10.56	10.30	13.50	79	5.02	1.63	12.39
Mn	77	650	863	779	972	7975	443.38	5.44	55.78
Fe	1600	19500	25031	25100	30700	154600	8012.21	1.19	18.26
As	0.6	8.40	12.57	12.10	15.30	163.80	7.35	6.06	87.84
Th	0.3	8.10	12.84	12.40	16.30	60	6.74	1.13	3.47
Ca	800	12600	35495	22300	43500	295200	36675.93	2.38	7.15
P	50	725	1641	1250	2330	16620	1250.29	2.20	12.13
La	0.9	31.30	42.02	41.50	51.50	162.3	18.26	0.73	2.46
Cr	0.25	9.20	17.63	14	20.60	808.40	26.16	16.77	386.10
Mg	700	3800	7347	5800	8100	104600	7479.31	5.06	35.80
Ti	5	750	1159	1240	1610	3270	618.70	-0.23	-0.65
Al	2100	26700	40584	41600	54050	94700	17036.21	0.01	-0.75
Na	20	690	3667	2600	5845	29490	3587.45	1.23	1.68
K	400	4700	14008	9500	18500	68200	12472.37	1.39	1.21

In order to facilitate the interpretation of results, varimax rotation was implemented, since it is an orthogonal rotation that minimises the number of variables that have high loadings on each factor, simplifying the transformed data matrix and assisting interpretation (Reimann et al., 2002). R-mode factor analysis was performed, and, the different factors obtained were studied and interpreted in accordance with their presumed, i.e. geogenic, anthropogenic or mixed (Tables 2, 3).

GeoDAS™ was used to produce interpolated geochemical maps of the normalised factor scores by means of the multifractal inverse distance weighted (MIDW) algorithm (Cheng et al., 1994; Lima et al., 2003). More detailed description of the method and their practical usage, one can find several publications (Cheng et al., 1994, 2000; Lima et al., 2003a, 2005; Albanese et al., 2007, 2015; Zuo et al., 2015; Zuo and Wang, 2016; Minolfi et al., 2018a, 2018b; Parsa et al., 2017; Rezza et al., 2018).

The concentration–area (C–A) fractal plot (Cheng et al., 1994, 2000; Cheng, 1999) was used to classify the interpolated factor score maps and capture the different spatial patterns (Figs. 4, 5). Computations (e.g. log-ratio transformations, regressions, and factor analysis) and graphical representations were implemented by the open source statistical software of R and CoDaPack (Comas-Cufí and Thió-Henestrosa, 2011).

2.1. Compositional discrimination analysis

2.1.1. Multivariate regression analysis

Based on elemental associations on robust biplots, multivariate regression analyses were performed using 6 variables (Zn, Pb, Cr, Ni, Fe, and Mn). Iron and Mn were involved as independent variables because both of them and especially their secondary oxy-hydroxides are susceptible of adsorbing Ni, Cr, Zn and Pb in soils (Salminen et al., 2005; Kabata-Pendias, 2011; Reimann et al., 2014). Additive log-ratio (alr) transformation was applied on the 6 variables using Ti as a common divisor, as it is one of the most stable and less mobile elements in soils (Kabata-Pendias, 2011). Additive log-ratio was chosen because it is not affected by the singularity problem in regression and correlation calculation (Aitchison, 1986; Pawlowsky-Glahn et al., 2015a), and it enabled us to reveal the proportional changes for each variable. The alr-transformation is based on the following equation:

$$y = alr(x) = [\ln \left(\frac{x_1}{x_D} \right), \ln \left(\frac{x_2}{x_D} \right), \dots, \ln \left(\frac{x_{D-1}}{x_D} \right)] \quad (1)$$

where x^1-x^{D-1} are the D-1 variables (all variables except the common divisor element) and x_D is the common divisor element (Aitchison, 1986).

In order to reduce the leveraging effects of outliers on the regression hyperplane, the least trimmed squares (LTS) fitting method was selected because it is highly robust and a fast algorithm for its computation is available (Roosseeuw and Van Driessen, 2002). In the LTS method the sum of some proportion of the smallest squared residuals is minimised. The proportion may vary from 50 to 100%, with 50% providing maximum protection against outliers. We used the compromise of 75% of the data. t-Test statistics and its related P-values were calculated at the 95% confidence level to test whether both of the independent variables (Fe/Ti and Mn/Ti) significantly contributed to the regression model or one of them could be dropped. The null hypothesis is that the corresponding model parameter equals 0, based on the extra sums of squares attributable to each variable if it is entered into the model last. F-test statistics and its P values were also calculated to analyse variance and to test whether significant relationship existed between each dependent (Zn/Ti, Pb/Ti, Cr/Ti and Ni/Ti) and independent variable.

2.1.2. Calculation and mapping of balances

Based on the multivariate regression analysis, sequential binary partition was implemented using the same 6 variables (Zn, Pb, Cr, Ni, Fe and Mn) by dividing them into specific groups of non-overlapping elements. Balances are particular ilr-coordinates (isometric-logratio) having orthonormal bases which can be interpreted in the D-1 (D: dimension) real space as ratios of elemental associations (Egozcue et al., 2003). Balances can be calculated using the following formula:

$$z_i = \sqrt{\frac{rs}{r+s}} \ln \frac{(\prod_{+} x_j)^{1/r}}{(\prod_{-} x_k)^{1/s}} \text{ for } i= 1, \dots, D-1 \quad (2)$$

where the products \prod_{+} and \prod_{-} only include parts coded with + and –, and r and s are the numbers of positive and negative signs (parts) in the i-th order partition, respectively (Egozcue and Pawlowsky-Glahn, 2005). Table 4 shows the sequential binary partitions of the calculated 5 balances (z1–z5).

The calculated 5 balances (z1–z5) have been normalised by using the linear min-max normalisation method (from 1 to 10), prior to the multifractal interpolation. The interpolated balance maps were classified based on the CA (Concentration-Area) fractal method (Cheng et al., 1994; Cheng, 1999) to display the spatial patterns and the proportional changes of elemental associations. In this study we focus and present the first 4 balance maps (z1–z4 or ilr1–ilr4) because these depict the spatial proportion of the investigated PTEs.

2.1.3. Bivariate regression on specific balances and their residual map

Finally, by taking into consideration the factor score and balance maps, and the linear relationship of the 4 PTEs with Fe and Mn, a bivariate regression analysis was performed using normalised ilr-2 (independent) and ilr-1 (dependent) variables. The least trimmed squares (LTS) fitting method was also selected here because it is highly robust and resistant against outliers (Rousseeuw and Van Driessen, 2002).

The obtained standardised residuals (residuals divided by standard error) were plotted against ilr-2 coordinates. We were particularly interested in large positive and negative standardised residuals which may indicate different sources of contamination or enrichment. The residuals plot was divided into several parts based on the classification of the ilr-2 and the residuals by means of Tukey's box-and-whiskers plot. The areas below the lower quartile (Q1) and above the upper quartile (Q3) of the standardised variables were called geogenic enrichment and anthropogenic contamination, respectively. These areas were further subdivided based on the lower and upper quartiles of the ilr-2 coordinates which indicate the limits of Cr-Ni and Zn-Pb enrichment or contamination, respectively.

The standardised residuals were interpolated by using the multifractal method and classified by means of the C-A plot (Concentration-Area) to reveal their spatial patterns. The residual map was overlain by those sampling points whose ilr-2 and standardised residual values were beyond or lower than their respective Q1 and Q3. In this way, we could see exactly which elemental associations a contamination or an enrichment is related to.

3. Results and discussion

3.1. Covariance and correlations between variables

Covariance and correlations between variables

The robust biplots explain 63.53% (PC1–PC2) and 57.84% (PC1–PC3) variability of the investigated 18 elements. They reveal several elemental associations related to the underlying geology and anthropogenic activities.

In Fig. 2A, Al–Ti–Th, Fe, Mn, As, La, and Co elemental association (A1) is characterised by the vicinity of their vertices and their rays point in the same direction. This association is likely to have a mixture of geogenic origin. Within A1, Al, Ti and Th vertices are superimposed among which Th has the highest vector length. In fact, A1 might be divided in two sub-groups, namely A_{1-1} (Al–Ti–Th), which is possibly related to pyroclastic deposits and A_{1-2} (Fe, Mn, Co, As, La), which is probably tied to the coprecipitation of Fe-, Mn-oxy-hydroxides in siliciclastic deposits (Buccianti et al., 2015; Zuzolo et al., 2017).

Sodium and K variables form clearly another association (A2) because their vertices are located close to each other and they point in the same direction. Both variables display high communalities (the length of the ray). The A2 may be associated with topsoils underlain by potassic and ultrapotassic rocks.

Copper and P elements belong to the A3 because their vertices overlap each other. Furthermore, A2 and A3 associations are geometrically symmetric with respect to the Mg vector meaning that A3 enhances a mixed behaviour of Cu and P which is likely related to volcanic materials and phosphate fertilisers.

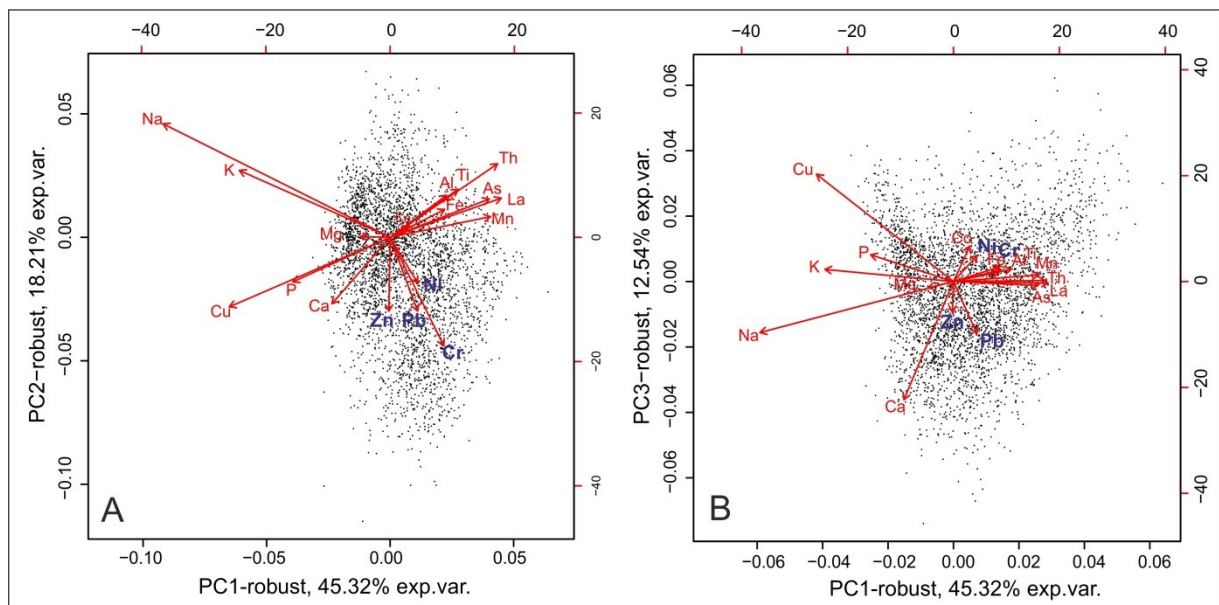


Fig. 2. Robust biplots for the first and second principal components (A) and for the first and third principal components (B) based on the 18 investigated elements.

Calcium and Mg vectors are symmetric around the A3 and form a separate group of A4. In fact, these two elements are dominants in carbonate rocks.

Zinc, Pb, Ni and Cr variables form another association (A5); their vertices lay closer to each other and point to the same direction. They may be related to anthropogenic sources like vehicular emission and industrial activities. However, Fig. 2B clearly shows the discrimination of the A5, where Pb and Zn vertices remain close to each other, but Ni and Cr changed their position and joined the elemental association of the A1. This could prove the affinity of Ni-Cr to the coprecipitation of Fe-, Mn-oxy-hydroxides in siliciclastic material. Compositional factor analysis was used to discern the correlations and interrelationships between variables, and their possible geochemical source patterns in the Campanian topsoils.

The total variance of the 18 elements is 77.26% in the four-factor model, which was chosen based on the break-point on the scree-plot of all factors. The 4 factors, named F1, F2, F3 and F4, account for 32.26%, 26.14%, 11.57% and 7.30% variability, respectively (Table 2).

Table 2. Varimax-rotated factors of logratio transformed (ilr) data; bold entries: loading values over absolute value of 0.5.

Variables	Factors				Communalities
	F1	F2	F3	F4	
Cu	-0.12	-0.77	0.23	-0.14	0.67
Pb	-0.28	0.10	0.09	-0.84	0.79
Zn	0.13	-0.21	-0.03	-0.83	0.75
Ni	0.90	-0.13	-0.28	0.09	0.91
Co	0.88	-0.10	0.01	0.37	0.92
Mn	0.69	0.41	0.04	0.05	0.65
Fe	0.79	0.31	0.20	0.27	0.83
As	0.01	0.72	0.25	-0.11	0.60
Th	-0.08	0.72	0.50	0.25	0.84
Ca	0.17	-0.17	-0.87	-0.06	0.83
P	-0.22	-0.72	0.00	0.03	0.57
La	-0.16	0.81	0.37	0.08	0.82
Cr	0.80	-0.08	-0.32	-0.20	0.78
Mg	0.20	-0.03	-0.82	0.14	0.73
Ti	-0.72	0.38	0.35	0.06	0.79
Al	-0.01	0.51	0.53	0.52	0.81
Na	-0.86	-0.24	0.17	0.14	0.84
K	-0.62	-0.46	0.30	0.50	0.94
Eigenvalues	5.81	4.71	2.08	1.31	
Total variance in %	32.26	26.14	11.57	7.30	
Cum. of total variance (%)	32.26	58.40	69.97	77.26	

Table 3. Explanation of the four-factor model extracted from the (ilr) logratio clr back-transformed data, which explains 77.26% of total variance.

Factors	% of variance explained	Association of variables	Interpretation
F1	32.26	(1) Ni, Co, Cr, Fe, Mn (2) Na, Ti, K	1. Pedogenic processes in flysch materials, with adsorption and co-precipitation of Fe and Mn oxy-hydroxides in oxidizing environment (Fig. 3A) 2. Potassic and ultrapotassic rocks and volcano-sedimentary deposits (Fig. 3A)
F2	26.14	(1) La, As, Th, Al (2) Cu, P	1. Pyroclastics rocks of the Mt. Roccamonfina volcano with high levels of incompatible elements (Fig. 3B) 2. Agricultural practices using phosphates fertilisers (Fig. 3B)
F3	11.57	(1) Al, Th (2) Ca, Mg	1. Pyroclastics deposits related to Mt. Roccamonfina (Fig. 4A) 2. reflects weathering processes of limestone and dolostone (Fig. 4A)
F4	7.30	(1) Al, K (2) Pb, Zn	1. Potassic pyroclastics composition of the Mt. Somma-Vesuvius (Fig. 4B) 2. Anthropogenic sources such high rate of fossil fuel combustion, industrial and vehicular emissions release (Fig. 4B)

Variables with loadings over the absolute value of 0.5 have been considered to describe the main composition of each factor. All variables hold communalities over 0.5 (50% of variability) meaning that the 4 factor models capture quite well the elemental interrelationships and their possible geogenic and/or anthropogenic sources (Table 3). The 18 elements of the four-factor model were separated by positive and negative loadings and sorted in descending order:

F1: Ni, Co, Cr, Fe, Mn, - (Na, Ti, K)

F2: La, As, Th, Al, - (Cu, P)

F3: Al, Th, - (Ca, Mg)

F4: Al, K, - (Pb, Zn)

The Fig. 3A shows interpolated map of factor scores (F1), ranging from -1.73 to 5.09. High factor scores (ranging from 2.88 to 5.09) are mapped in the eastern and south-western parts of the study area where siliciclastic deposits prevail (Fig. 3A). Pedogenic processes in flysch materials (fine-size) are mostly characterised by adsorption and coprecipitation of Co, Cr and Ni with Fe and Mn oxy-hydroxides in oxidizing environment (Postma, 1985). Low factor scores (ranging from -1.73 to -0.75) are revealed close to large volcanic complexes like the Phlegraean Fields, Mt. Somma-Vesuvius and Mt. Roccamonfina.

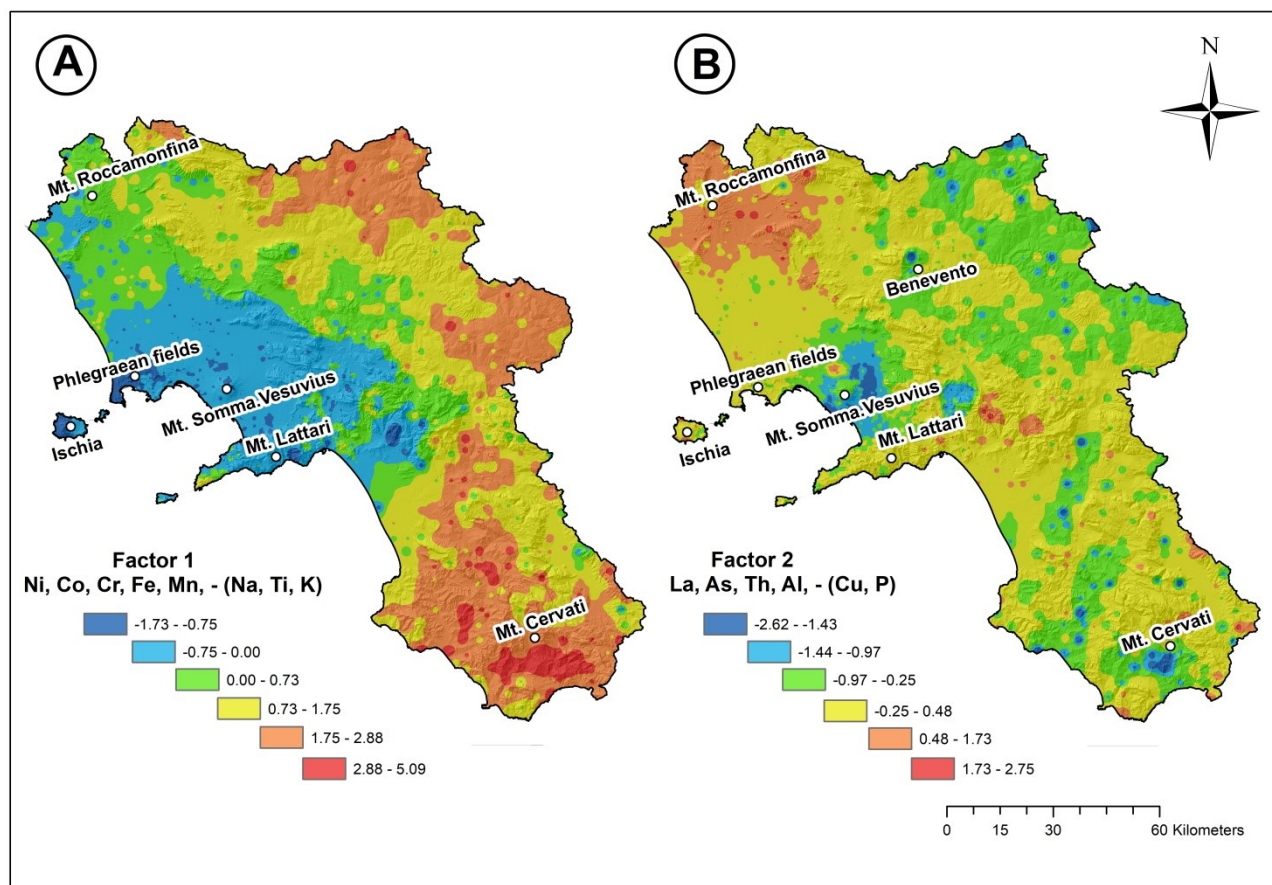


Fig. 3. Interpolated factor score maps of the Factor 1 (A) and Factor 2 (B).

Negative scores are given by the antithetic elemental association including Ti, Na, K. These elements are likely related to the potassic and ultrapotassic rocks and volcano-sedimentary deposits from major sector collapse of volcanoes in the survey area (Scheib et al., 2014; De Vivo et al., 2016).

High factor scores of the F2 range from 1.73 to 2.75 and are related to the positive correlation between Al, La, As, and Th in soils of the Mt. Roccamonfina and surroundings (Fig. 3B). Pyroclastic rocks of the Mt. Roccamonfina volcano are characterised by high levels of incompatible trace elements with an orogenic signature having troughs at Ba, Ta, Nb, and Ti, and peaks at Cs, K, Th, U, and Pb (Conticelli et al., 2009). Low factor scores (ranging from -2.62 to -1.43) can be observed around the Mt. Somma-Vesuvius, Benevento (Vitaliano municipality), and in the Mt. Cervati. Large agricultural fields like vineyards and orchards are located in these areas. Copper and P are likely related to agricultural practices using phosphate fertilisers (Thiombane et al., 2018a).

The F3 factor score map (Fig. 4A) shows (particularly for Al and Th elemental association) elevated values (ranging from 1.26 to 2.39) close to the Mt. Roccamonfina. The Mt. Roccamonfina volcano is characterised by pyroclastic rocks with high level of trace elements such Th, U, and La (Conticelli et al., 2009). But relatively high factor scores (ranging from 0.51 to 1.26) can be observed around the Phlegraean fields and the Mt. Somma-Vesuvius, and along the Apennines which were related to the occurrences of Campanian Ignimbrites (De Vivo et al., 2010). In contrast, low factor scores (Ca and Mg elements) are mapped exactly in correspondence with the three large river plains of the Campania Region, namely Volturno, Sele and Telesina. Furthermore, these areas are well drained by rivers and tributaries which spring from the carbonate rocks of the Apennine chain. Indeed, Ca and Mg in these areas are mostly associated with the weathering processes of limestone and dolostone (Peccerillo, 2005; Albanese et al., 2007).

The F4 score map (Fig. 4B) shows high positive values (ranging from 1.29 to 2.77) around the Mt. Somma-Vesuvius, where Al and K are the most congruent elements with the potassic pyroclastics composition (Peccerillo, 2001; De Vivo et al., 2010). Low factor scores (particularly Pb and Zn elemental association) are mapped exclusively on topsoils of the most urbanised areas (Naples, Salerno and Castelvoturno) of the study area. Lead and Zn may be tied to anthropogenic sources such high rate of fossil fuel combustion, industrial and vehicular emissions release.

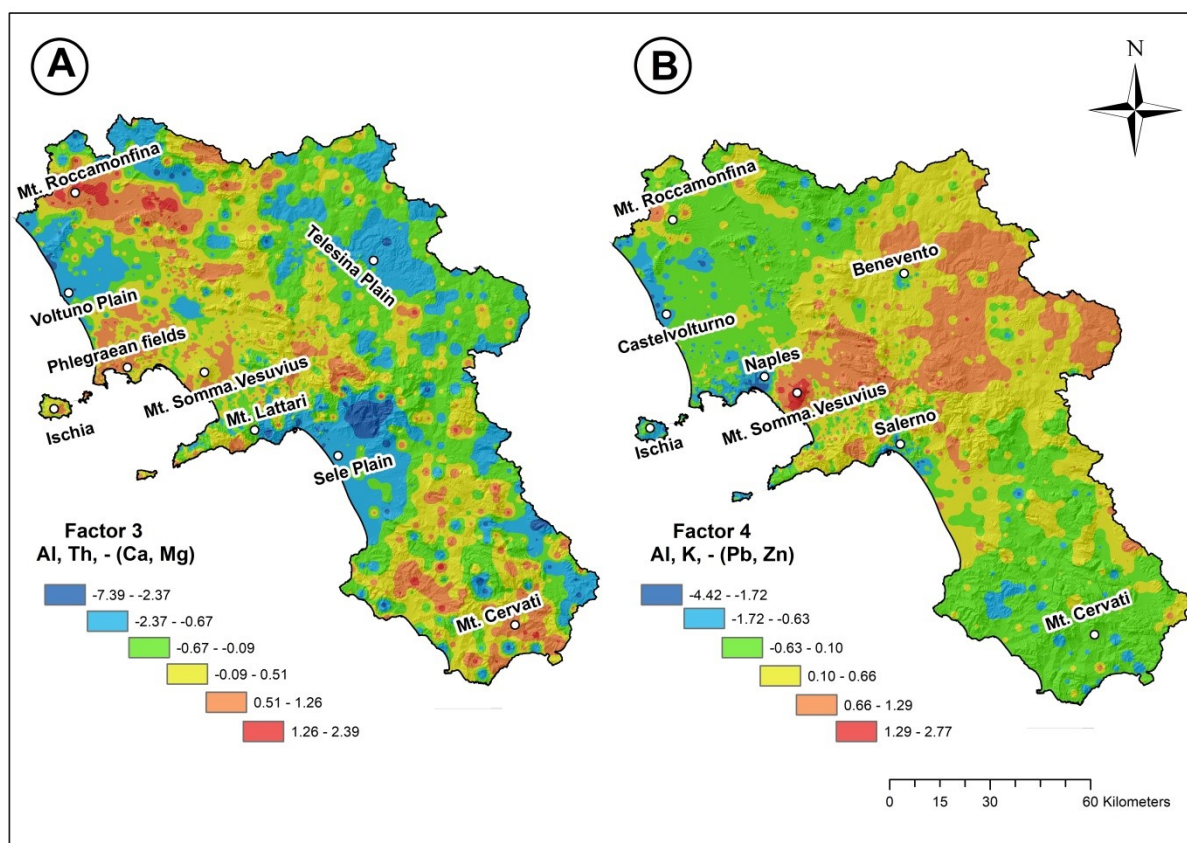


Fig. 4. Interpolated factor score maps of the Factor 3 (A) and Factor 4 (B).

3.2. Compositional discrimination analysis

In this session, we present the results of compositional data analysis used to discriminate the main sources of the investigated potentially toxic elements (Zn, Pb, Cr and Ni).

3.2.1. Multivariate regression analysis

Multivariate regression analysis was implemented by using the alr-transformed coordinates of 6 elements (Zn, Pb, Cr, Ni, Fe and Mn) in order to examine their relationships and correlations (Fig. 5).

The multivariate regression analysis revealed strong positive correlation ($R^2 > 54.53\%$) and statistically significant (F-test statistics, at 95% confidence level) linear relationship between each dependent (Zn/Ti, Pb/Ti, Cr/Ti, Ni/Ti) and independent (Fe/Ti, Mn/Ti) alr-coordinates (Fig. 5). Based on the t-test statistics and its associated P-values at the 95% confidence level, both of the independent variables significantly contribute to the predicted model results, hence neither of them can be dropped.

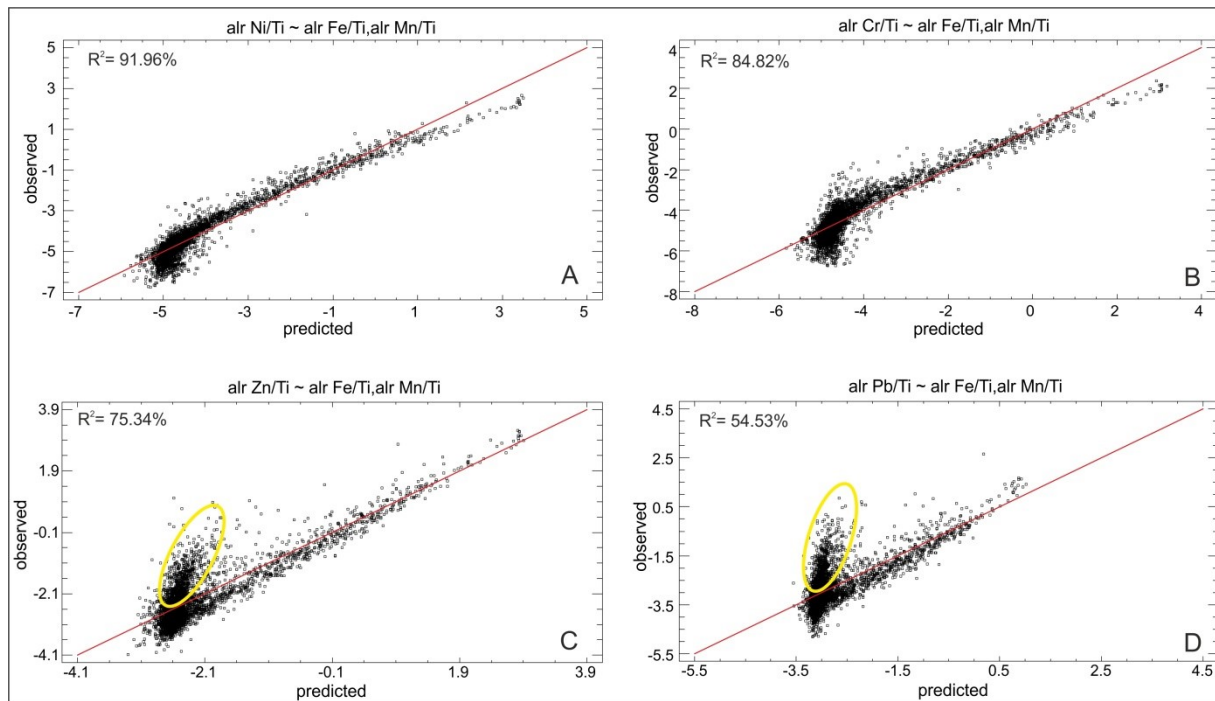


Fig. 5. Multivariate regression analysis of the alr-transformed Ni, Cr, Zn and Pb (dependent variable) against Fe, Mn (independent variables). Note the good correspondence and strong linear relationships between predicted and observed values. The high positive correlation (R^2) decreases from Ni to Pb. In addition, the yellow ellipses at Zn and Pb mark deviation tails from linearity suggesting different processes.

It should be noticed that the positive correlation decreased continuously from Ni/Ti (R^2 : 91.96%), Cr/Ti (84.82%) to Zn/Ti (75.34%) and Pb/Ti (54.53%) (Fig. 5). In addition, interesting deviation tails from the linear trends were discerned in case of Zn/Ti and Pb/Ti which suggest different processes (Figs. 5C, D). The deviations can be observed at low alr-Fe/Ti and alr-Mn/Ti which may indicate other sources (anthropogenic?) of Zn and Pb besides the accumulation of Fe-, and Mn-oxy- hydroxides.

3.2.2. Balance (ilr) maps

Based on the multivariate regression analysis, the same 6 elements (Zn, Pb, Cr, Ni, Fe and Mn) have been chosen to perform sequential binary partition and obtain balances (specific ilr-coordinates) (Table 4).

Table 4. Sequential binary partition table of the 6 investigated variables and the obtained balances (z1–z5). Parts coded with + and – are the elemental associations involved in the calculation of the i-th order partition, respectively.

	Zn	Pb	Cr	Ni	Fe	Mn
z1 (ilr-1)	+	+	+	+	-	-
z2 (ilr-2)	+	+	-	-		
z3 (ilr-3)	+	-				
z4 (ilr-4)			+	-		
z5 (ilr-5)					+	-

3.2.3. Ilr-1 map (ZnPbCrNi/FeMn)

The first balance (ilr1) map reveals higher proportion of the PTEs around large urban and industrial areas (e.g. Naples, Salerno, and Castelvoturno). These cities are located all along the coastline where the highest population density and the majority of industrial settlements and traffic are concentrated (Fig. 6A).

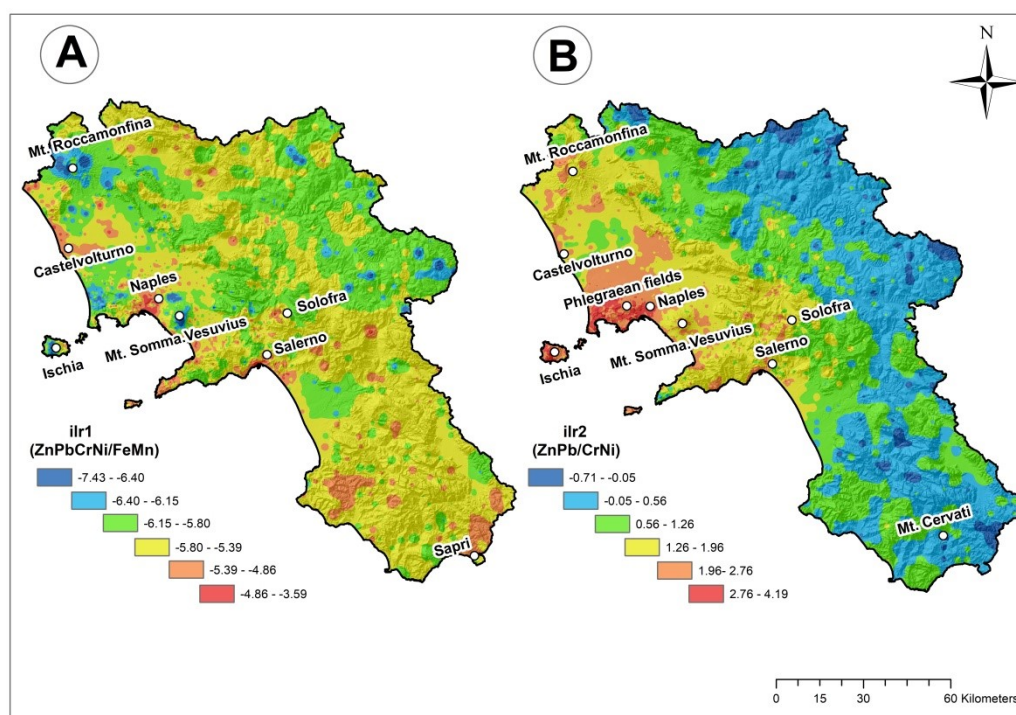


Fig. 6. (A) The interpolated ilr-1 map. Note the higher proportion of the PTEs (with red colour) around Naples and Salerno. (B) The interpolated ilr-2 map. The higher spatial abundance of Cr and Ni (with blue colour) is clearly related to the eastern and southern siliciclastic zones.

In contrast, the higher abundance of Fe and Mn can be observed close to large volcanic edifices (e.g. Mt. Roccamonfina, Mt. Somma-Vesuvius) and their surroundings. Some high Fe and Mn proportion patches are also displayed in the eastern siliciclastic zone indicating that the behaviour and accumulation of these elements are mainly influenced by the underlying bedrocks (Fig. 6A).

3.2.4. Ilr-2 map (ZnPb/CrNi)

The second balance map is based on the subdivision of PTEs taking into consideration their separate associations on the robust biplot (PC1–PC3) and their factor loadings (Fig. 2, Table 2). The higher proportion of Zn and Pb is associated with urban and industrial areas (e.g. Naples and Salerno) but also reaches higher abundance close to large volcanic complexes like Phlegraean Fields and Ischia (Fig. 6B). In contrast, the higher proportion of Cr and Ni can be observed exclusively in the eastern and southern part of the study area, far away from most urbanised areas (Fig. 6B). These elements have higher abundance in topsoils over siliciclastic deposits (e.g. flysch sediments) in the Apennine Mts; hence they are much related to specific lithology (e.g. flysch, claystone) and related pedogenic processes (Figs. 1A, 6B). According to Buccianti et al. (2015), the Ni and Cr abundance in Campania is highly influenced by dispersion mechanisms that characterize sediment material particularly flysch deposits. It is interesting to see the NW-SE trending edge in the middle of the study area which separates the two different proportion zones and coincides roughly with the boundary of siliciclastic and carbonate outcrops (Fig. 6B).

3.2.5. Ilr-3 map (Zn/Pb)

The third balance map is based on the ratio between Zn and Pb and sheds light on their spatial abundance (Fig. 7A). The map unveils that the higher Zn proportion zones are mainly related to underlying siliciclastic deposits in the eastern and southern part of the study area, far away from urbanised regions (Fig. 7A). According to Minolfi et al. (2018b), the behaviour and accumulation of Zn is mainly influenced by the underlying bedrocks in the Campania Region. Buccianti et al. (2015) revealed that Zn has one the lowest clr-variances (central logratio) meaning that it maintains similar abundance in the Campanian topsoils.

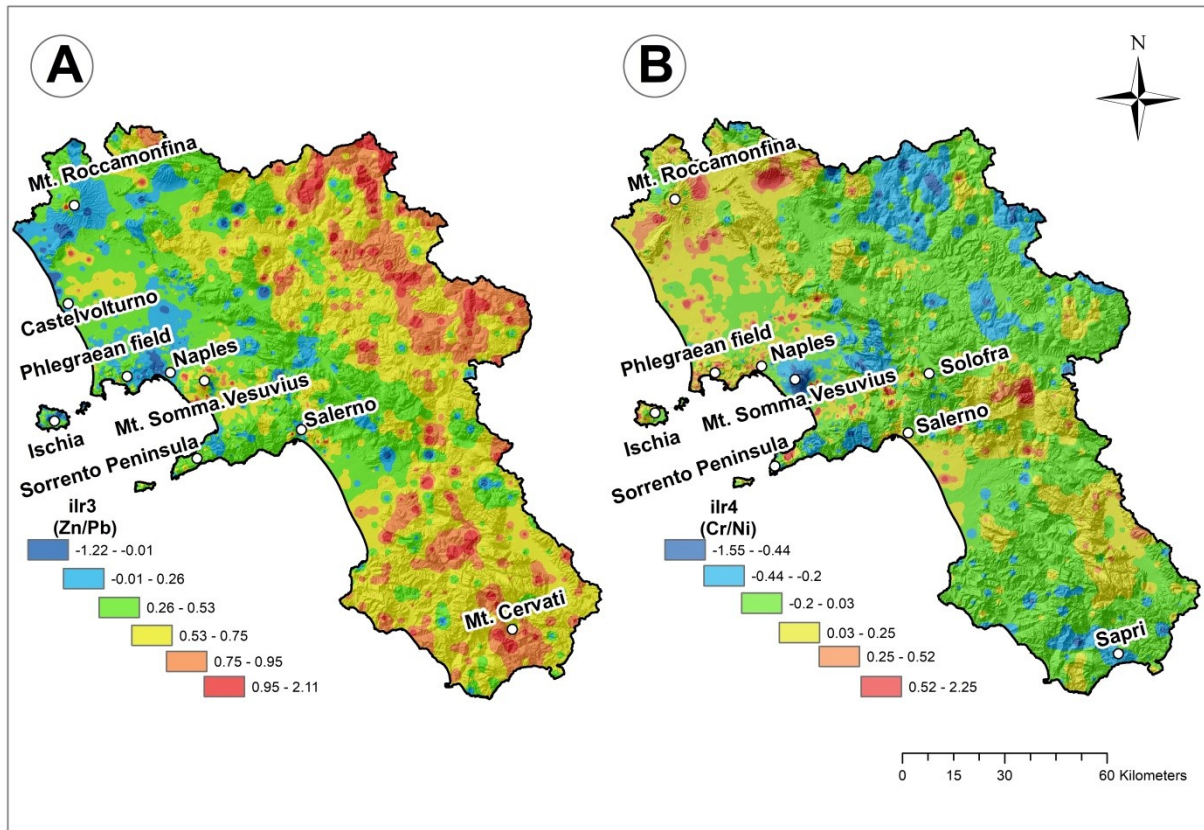


Fig. 7. (A) The interpolated ilr-3 map. Note the higher proportion of Zn (with red colour) in the eastern and southern siliciclastic zones. Pb proportion is the highest around Naples. (B) The interpolated ilr-4 map. The higher spatial abundance of Ni is related to the Mt. Somma Vesuvius and the eastern siliciclastic zone. The higher Cr proportion can be observed around Solofra, for example.

The higher Pb proportion zones are located in the urbanised area of Naples and some parts of Ischia and Sorrento Peninsula where they are tied to intensive traffic (Fig. 7A). The highest concentration of Pb (up to 425 mg/kg) in stream sediment was also described from Naples and that was explained by heavy traffic and human activities (Lima et al., 2003a; De Vivo et al., 2016; Minolfi et al., 2018b). However, the higher Pb proportion zone can also be revealed around Mt. Roccamonfina, where the average Pb concentration from phonolite and phonotheprite was 129 mg/kg (Peccerillo, 2005). This concentration is two times higher than the counterpart of Campanian volcanic rocks (64 mg/kg) (Paone et al., 2001); hence the Pb enrichment in the Roccamonfina area is more likely related to the underlying volcanic rocks.

3.2.6. Ilr-4 map (Cr/Ni)

The fourth balance map demonstrates the relative abundance of Cr and Ni and their spatial distribution (Fig. 7B). The map shows that their proportion is similar in the vast majority of the study area (ilr-4: -0.2 – 0.25). One of the higher Cr proportion areas is related to the Sarno Basin (e.g. Solofra district) where several factories (e.g. tannery industry) operate contributing to the abundance of Cr (Fig. 7B).

Some high Cr proportion patches are observed around the Mt. Roccamonfina and above some carbonate massifs (e.g. Mt. Matese) where the most advanced pedogenetic processes (e.g. rubification, eluviation) were demonstrated (Petrik et al., 2018a). The trivalent chromium (Cr^{3+}) behaves much like Al^{3+} and Fe^{3+} and tends to accumulate with secondary oxy-hydroxides in matured soils especially under alkaline environment (Salminen et al., 2005; Reimann et al., 2014).

Nickel reaches higher proportion compared to Cr in the eastern part of the study area in topsoils over claystone and siltstone (Fig. 7B). The lowest mobility and variance of Ni concentrations were described here by Petrik et al. (2017). The higher abundance of Ni close to the Mt. Somma-Vesuvius may be related to underlying volcanic rocks.

3.2.7. Bivariate regression analysis and residual map

Based on the factor score and balance maps and the strong positive correlation (R^2 : 54.53–91.96%) of the 4 PTEs with Fe and Mn, we performed bivariate regression analysis using the normalised ilr-2 (independent) and ilr-1 (dependent) variables to discriminate and map the anthropogenic and geogenic parts of the contamination (Fig. 8).

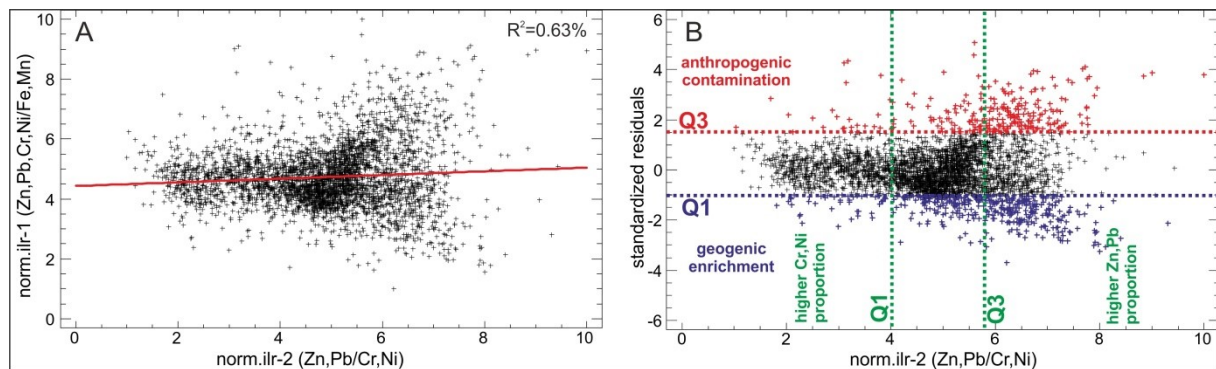


Fig. 8. (A) Bivariate regression analysis between normalised ilr-2 (independent) and ilr-1 (dependent) variables. (B) Standardised residuals against normalised ilr-2. The plot was divided into more parts based on the classification of ilr-2 and residuals by means of Tukey's box-and-whiskers plot. Areas outside of lower (Q1) and upper quartiles (Q3) belong to geogenic enrichment and anthropogenic contamination of different elemental associations.

We were particularly interested in standardised residuals which were plotted against normalised ilr-2 coordinates (Fig. 8B). The large positive (> 1.55) standardised residuals reveal low Fe/Mn proportion (ilr1), hence mainly indicating anthropogenic contamination. In contrast, large negative (< -1.04) standardised residuals (high Fe/Mn proportion) represent geogenic contamination (Fig. 8B).

The standardised residuals of the bivariate regression were used to make residual map which was classified by means of the C-A (Concentration-Area) plot (Fig. 9). The residual map was overlain by

those sampling points which are related to different sources of contamination or enrichment based on the standardised residuals plot (Figs. 8B, 9). The large positive residuals (> 1.55 , with red colour) all indicate anthropogenic contamination of the 4 PTEs (Figs. 8B, 9). The densely populated, large urban and industrial areas along the coastline (e.g. Naples, Salerno and Castelvoturno) are all displayed as highly contaminated regions mainly due to the heavy traffic (Pb anomaly) and various heavy industries (e.g. production of alloys: Zn). Some scattered, high contamination patches in the Sarno Basin are likely related to intensive agriculture using phosphate fertilisers (Zn, Cr, Ni anomaly).

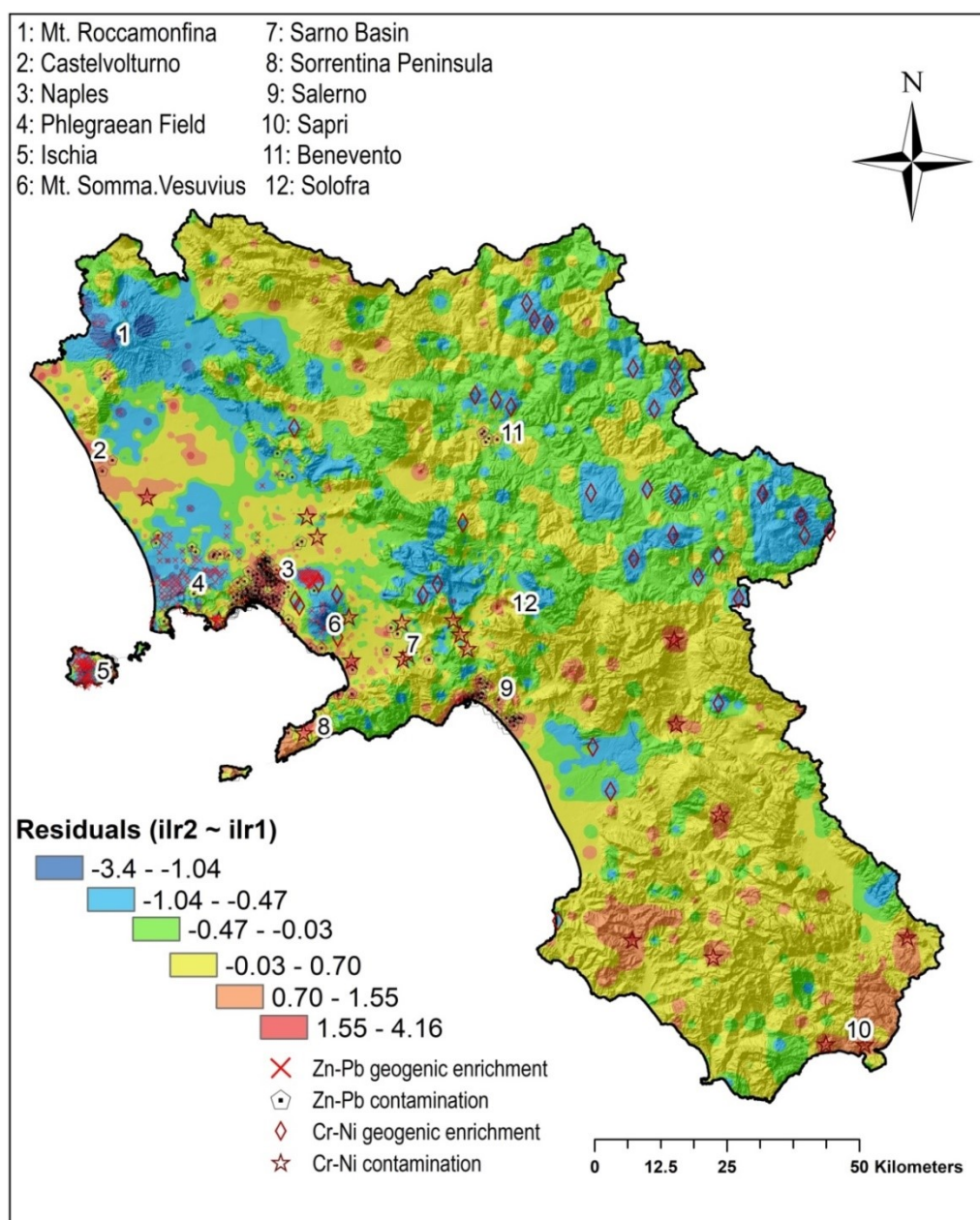


Fig. 9. Classified standardised residual map indicating the contamination and enrichment areas. The map is overlain by those sampling points which are related to anthropogenic contamination or geogenic enrichment of the respective 4 PTEs.

Contamination in the Solofra industrial district is linked to traffic emission (Pb), alloy production (Zn and Ni) and tannery (Cr and Ni) industries (Fig. 9). The southern urbanised region around Sapri is also displayed as contaminated area, which was associated with large urban areas (Pb) and smelting industries (Cr, Ni) (Thiombane et al., 2018a).

In contrast, the large negative residuals (<-1.04 with dark blue) indicate the geogenic source of the four investigated PTEs. The large volcanic complexes (e.g. Mt. Somma-Vesuvius, Phlegraean Fields, Mt. Roccamonfina) are all characterised by negative residuals demonstrating the main influence of underlying geology (e.g. different volcanic rocks) on the proportion of Zn and Pb (Fig. 9). Some negative residuals can be discerned in the eastern part of the study area where the anomaly of Cr and Ni may be influenced not only by the presence of Fe-, and Mn-oxy-hydroxide but also the organic matter and clay content (Fig. 9). According to Buccianti et al. (2015), the Ni and Cr distribution are strongly dependent on the various lithology (e.g. clay content) and grain size of the siliciclastic zone in the Apennine Mts.

4. Conclusion

This study demonstrates a comprehensive discrimination analysis carried out on 4 PTEs (Zn, Pb, Cr and Ni) of the Campanian high-density topsoil data set. A workflow of compositional data analysis was implemented ranging from robust biplot and factor score maps to multivariate and bivariate regression analyses to discriminate the possible sources of contamination/enrichment.

The study represents a new discrimination index based on standardised residuals of the bivariate regression of the $ilr-2$ against $ilr-1$ coordinates which enabled us to identify and separate the anthropogenic contamination from the geogenic enrichment of the respective 4 PTEs. The large urban and industrial areas (e.g. Naples, Salerno) along the coastline are mainly contaminated by Pb and Zn due to heavy traffic and alloy production. Some Cr and Ni contamination was discerned in the Sarno Basin due to the release of waste rich in Cr-Ni deriving from tannery industries (e.g. Solofra). The large volcanic complexes (e.g. Mt. Somma-Vesuvius, Mt. Roccamonfina) are all characterised by geogenic enrichment of Zn and Pb. In contrast, Cr and Ni-enrichment is mainly related to the siliciclastic deposits where their proportion is not only influenced by Fe and Mn but also organic matter, clay content and dispersion mechanism.

Section 2.4

Exploratory analysis of multi-element geochemical patterns in soil from the Sarno River Basin (Campania region, southern Italy) through compositional data analysis (CODA)

This section has been published in:
Journal of Geochemical Exploration
Volume 195, December 2018, Pages 110-120

Exploratory analysis of multi-element geochemical patterns in soil from the Sarno River Basin (Campania region, southern Italy) through compositional data analysis (CODA)

Abstract

The Sarno River Basin (south-west Italy), nestled between the Somma–Vesuvius volcanic complex and the limestone formations of the Campania–Apennine Chain, is one of the most polluted river basins in Europe due to widespread industrialization and intensive agriculture. Water from the Sarno River, which is heavily contaminated by the discharge of human and industrial wastes, is partially used for irrigation on the agricultural fields surrounding it. We apply compositional data analysis to 319 soil samples, collected during two field campaigns along the river course and throughout the basin, to determine the concentration and possible origin (anthropogenic and/or geogenic) of the elemental anomalies, including potentially toxic elements (PTEs).

The concentrations of 53 elements were determined using ICP-MS and, subsequently, log-transformed. Using hierarchical clustering, clr-biplot and a principal factor analysis, the variability and the correlations between a subset of extracted variables (26 elements) were identified. Factor score interpolated maps were then generated using both lognormal data (NDR) and clr-transformed data to better visualize the distribution and potential sources of the patterns in the Sarno Basin.

The underlying geology substrata appear to be associated with raised levels of Na, K, P, Rb, Ba, V, Co, B, Zr, and Li, due to the presence of pyroclastic rocks from Mt. Somma–Vesuvius. Similarly, elevated Pb, Zn, Cd, and Hg concentrations are most likely related to the geological and anthropogenic sources, the underlying volcanic rocks, and contamination from fossil fuel combustion associated with nearby urban centers. Interpolated factor score maps and the clr-biplot show a clear correlation between Ni and Cr in samples taken along the Sarno River, and Ca and Mg near the Solofra district. After considering nearby anthropogenic sources, the Ni and Cr are PTEs most likely originating from the Solofra tannery industry, while Ca and Mg correlate to the underlying limestone-rich soils of the area.

This study shows the applicability of the log-ratio transformations to these studies, as they clearly show relationships and dependencies between elements which can be lost when univariate and classical multivariate analyses are employed on raw and lognormal data.

1. Introduction

Environmental geochemistry investigates the impact of natural geochemical processes and anthropogenic activities on natural systems (e.g., rivers, lakes, soils, or forests), and on human health. These activities may lead to issues of ongoing anthropogenic global environmental change such as acid rain, ocean acidification, toxic metal pollution, or water contamination, and may result in contamination

of an area by potentially toxic elements (PTEs). A useful tool in determining the level and distribution of PTE contamination, and thereby identify areas where there may be a risk to the local ecology and to humans, is an environmental risk assessment (Aral, 2010; Health Canada, 2010a, 2010b; DEFRA, 2011).

The Sarno Basin, considered to be one of the most polluted areas of Europe, is contaminated by PTEs from both anthropogenic sources including several different industries (Arienzo et al., 2001; Cicchella et al., 2014), as well as from the underlying geological substrata. Statistical studies may be employed to define the characteristics of the PTEs in the Sarno Basin, including the definition of the background concentrations of these elements, the relationships between different chemical elements, and, in some cases, the determination of the extent of anthropogenic effects on the contamination status of the PTEs in the basin. It is understood that classical statistical methods based on Euclidean distance, and particularly those based on covariance structure, are not applicable to geochemical concentrations (Aitchison, 1986). This is mainly due to the positive character of geochemical data, and to the fact that they carry only information about the relative abundance of each individual component. To treat these compositional data using multivariate analysis, we must assume that the sample space is R^D , which can properly reconcile the correlation between elements. This was first recognized by Pearson (1897), who introduced the concept of spurious correlation for coefficients computed from raw compositions, which has since been studied by many authors, in particular by Chayes (1960), who tried to relate these issues to the singularity of the covariance matrix. Aitchison (1986) was the first to put forward a comprehensive alternative approach to classical statistics in order to avoid this problem, defining a new distance more suitable to compare compositions than the Euclidean distance. The most common approach to deal with concentration variables is now the use of log-ratio transformation (Aitchison, 1986; Aitchison et al., 2000; Aitchison and Egozcue, 2005) to express the compositions in terms of log-ratio coordinates. Three major transformations have been developed: additive log-ratio (alr), centered log-ratio (clr), both by Aitchison (1986), and the isometric log-ratio (ilr) by Egozcue et al. (2003). This study uses the clr-transformation method, as the clr-coordinates are most appropriate to interpret the distances between elements (Palarea-Albaladejo et al., 2012; Thiombane et al., 2018a), and to perform a dimensionality reduction through use of a biplot (Aitchison and Greenacre, 2002). Clr transformation was applied to the data in conjunction with clr-biplot compositional statistical transformation and Factor Analysis (FA), to distinguish the origin of the pollution in the Sarno River Basin. The main objectives of this multivariate investigation were:

- to determine the geochemical anomalies and sources of element patterns in the Sarno Basin.
- to distinguish the main vector and element associations into the environment.

- to use log transformation methods (lognormal and clr log-transformation) on the data to distinguish the potential anthropogenic or geogenic origin of the PTEs in the basin by applying clr-biplot and FA.
- to assess how the main lithological chemistry and human activities affect the Sarno Basin by using interpolated factors scores maps.

To facilitate better interpretation of the distribution patterns and the correlation between elements, the results of the clr transformed data were compared with the classical log transformed data by means of interpolated maps with ArcGIS software. Log-ratio transformations and clr-biplot were made using CoDaPack software (Comas-Cufí and Thió-Henestrosa, 2011), whereas the R Software was used to determine the correlation matrix and perform the FA.

2. Study area

Covering an area of 500 km², the Sarno River Basin is located in the Campania Region (southern Italy). The basin is bounded by the pyroclastic deposits of the Somma–Vesuvius volcanic complex to the northwest, the Tyrrhenian Sea to the west, the Picentini and Sarno Mountains to the east, and by Lattari Mountains and Sorrento Peninsula limestone reliefs to the south-west (Fig. 1A). Carbonate rocks belonging to the Campano–Lucanian and Abruzzese–Campanian carbonate platforms crop out in the Sarno River Basin including limestone outcrops around the Sarno and Lattari Mountains, Triassic dolomite, Lower Jurassic–Cretaceous limestone and dolomite limestone, and fractured and karstified Cretaceous limestone (De Pippo et al., 2006). Volcanic tuff, lapilli, and ash deposits from activity of the Somma–Vesuvius volcanic complex generally cover the calcareous-dolomitic rock of the Sarno Mountains (Fig. 1A). It is generally understood that the main soil characteristics of the study area are the result of geological phenomena (weathering, soil erosion) of the lithological substrate in and surrounding the Sarno Basin.

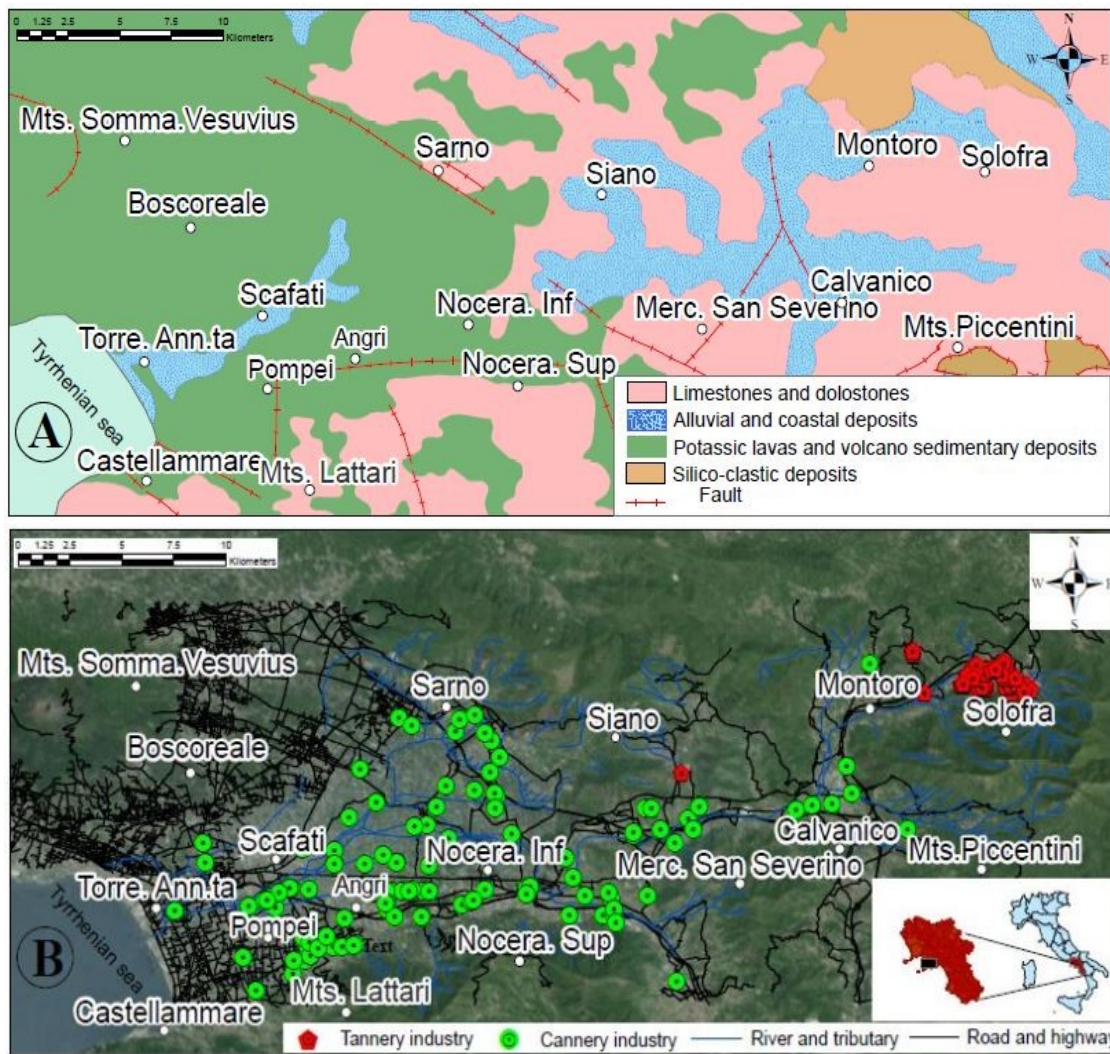


Fig. 1. The Sarno River Basin; (A) Simplified geological map of the Sarno Basin; (B) Map of land use and industrial activities.

The Sarno River flows approximately 24 km through the basin, and is considered the most polluted river in Europe (Adamo et al., 2008; Albanese et al., 2013). Untreated domestic and agricultural effluent and solid and liquid industrial wastes are discharged directly into the river. It springs from the slopes of Mt. Sarno near the eponymous town, and empties into the Gulf of Naples (Tyrrhenian Sea) near Rovigliano Rock, between Castellammare di Stabia and Torre Annunziata (Arienzo et al., 2001). The river system traverses the Campanian provinces of Salerno, Avellino, and Naples, and collects water from the Solofrana and Cavaioia tributaries, as well as discharge from the municipalities of Montoro, Scafati and Mercato San Severino.

The population of the Sarno Basin is approximately 1,400,520 inhabitants, with an urban average density of about 1859 inhabitants per km² and a peak density along the coastal and some inland areas of over 2200 inhabitants per km² (ISTAT, 2013). The Sarno Basin is highly industrialized with >160 tanneries located mostly in the Solofra area (Arienzo et al., 2001), canneries distributed along the length

of the river, and the industrial area of Castellammare containing the Novartis pharmaceutical factories, and medium-scale industrial facilities producing paints, ceramics, and packaging food (Fig. 1B). The situation is aggravated by pervasive discharge of untreated agricultural and industrial waste directly into the river. The Sarno Valley is also home to very intensive agricultural activities, consisting mainly of tomato field horticulture in San Marzano, extensive orchards between Pompeii and Scafati, vineyards in the Boscoreale area, and chestnuts, greenhouse horticulture, floriculture, and crop production throughout the basin.

3. Materials and methods

3.1. Soil sampling and preparation

Soil sampling was carried out during two campaigns. During the first (known as the SAR–SAR campaign), topsoil samples were collected from a depth of 0–20 cm from the surface, at 36 sampling points of irrigated and flood-plain soil on average every 0.8 km along the main course of both the Sarno and Solofrana rivers in 2010. The second campaign (known as the SAR–SOB campaign) collected 283 soil samples from throughout the Sarno River Basin, with an average sampling density of approximately one sample per 2 km². In total, 319 samples were collected during the two campaigns (Fig. 2), following internationally recognized sampling methods (Plant et al., 1996; Salminen et al., 1998).

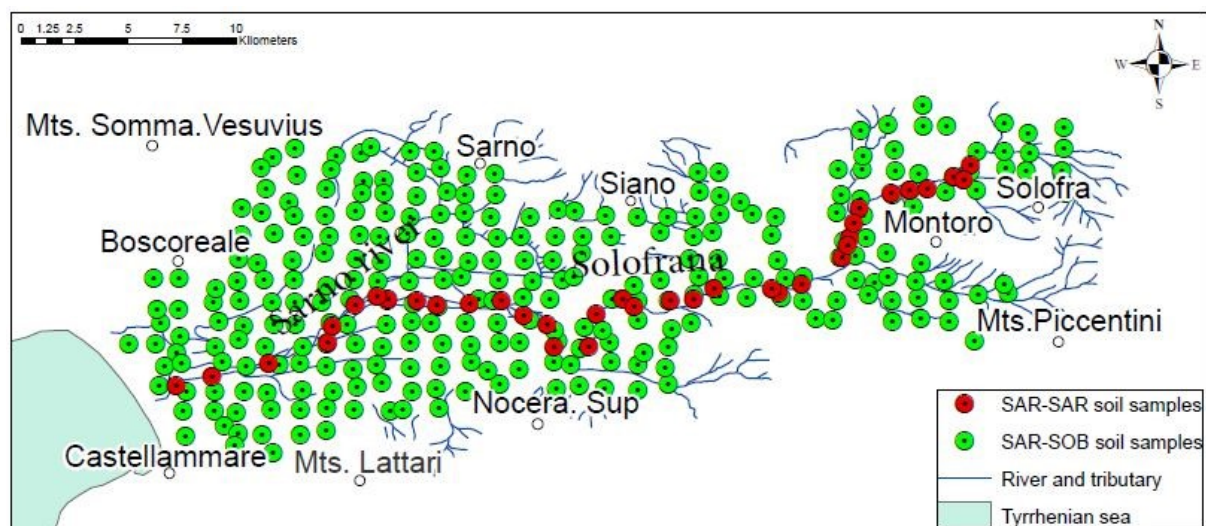


Fig. 2. Map of the study area with the location of soils sampling sites: SAR–SAR represents samples sites in along the length of the Sarno River; SAR–SOB samples sites are distributed across the Sarno Basin.

Each soil sample was collected as composites from five pits within an area of 100m², and consisted of approximately 3 kg of soil, collected between 0 and 20 cm below the surface, stored in inert plastic

bags. All samples were air-dried to prevent the volatilization of Hg, and sieved to collect 30 g of the <2mm fraction for chemical analysis.

3.2. Chemical analysis

Analyses were carried out at Acme Analytical Laboratories Ltd (now Bureau Veritas, Vancouver, Canada), an ISO 9002 accredited laboratory, within a 20 day time frame from when the samples were received, to the final delivery of analytical results. The samples were analyzed after an aqua regia extraction, by a combination of inductively coupled plasma atomic emission (ICP-AES) and inductively coupled plasma mass spectrometry (ICP/MS). The analytical procedures are highlighted in the section 2.1, where with the same protocol has been developed. Table 1 shows the results of the 40 considered elements.

Table 1. Detection limits, accuracy and precision of the applied analytical method.

Element	Unit	detection limit	accuracy $\Delta(\%)$	Precision (%)
Al	%	0.01	0.4	1.2
Ca	%	0.01	2.5	1.2
Fe	%	0.01	0.7	1.8
K	%	0.01	2.5	3.8
Mg	%	0.01	1.5	1.4
Na	%	0.001	3.3	2.7
P	%	0.001	1.2	3.8
S	%	0.02	19.4	11.5
Ti	%	0.001	2.5	5.7
Ag	mgKg ⁻¹	0.002	6.5	10.9
As	mgKg ⁻¹	0.1	1.3	2.2
Au	mgKg ⁻¹	0.0002	8.2	21.9
B	mgKg ⁻¹	1	0.4	8.2
Ba	mgKg ⁻¹	0.5	0.3	1.7
Be	mgKg ⁻¹	0.1	0.7	2.3
Bi	mgKg ⁻¹	0.02	1.2	3.5
Cd	mgKg ⁻¹	0.01	1.4	5.1
Co	mgKg ⁻¹	0.1	2.3	3.2
Cr	mgKg ⁻¹	0.5	2.5	1.2
Cu	mgKg ⁻¹	0.01	1.6	3.7
Ga	mgKg ⁻¹	0.1	3.3	4.2
Hg	mgKg ⁻¹	0.005	2.4	6.3
La	mgKg ⁻¹	0.5	3.5	2.9
Mn	mgKg ⁻¹	1	1.5	2.9
Mo	mgKg ⁻¹	0.01	1.4	1.1
Ni	mgKg ⁻¹	0.1	0.8	2.7
Pb	mgKg ⁻¹	0.01	1.4	1.8
Sb	mgKg ⁻¹	0.02	1.3	2.3
Sc	mgKg ⁻¹	0.1	0.8	4.4
Se	mgKg ⁻¹	0.1	0.9	22
Sn	mgKg ⁻¹	0.1	0.5	3.4
Sr	mgKg ⁻¹	0.5	4.3	2.5
Te	mgKg ⁻¹	0.02	3.9	6.4
Th	mgKg ⁻¹	0.1	4.1	3.1
Tl	mgKg ⁻¹	0.02	2.2	4.1
U	mgKg ⁻¹	0.1	1.6	3.7
V	mgKg ⁻¹	2	2.3	3.2
W	mgKg ⁻¹	0.2	3.7	5.1
Zn	mgKg ⁻¹	0.1	1.8	2.2

3.3. Statistics and compositional data analysis (CoDA)

A composition is defined as a sample space of the regular unit D simplex, S^D , that is a vector of D positive components summing up to a given constant k , set typically equal to 1 (proportions), 100 (percentages), or 10^6 (ppm) by closure. It relates parts of some whole that carry relative information (ratios of components) whose sample space is the simplex.

By analogy to the lognormal approach, Aitchison (1986) projected the sample space of compositional data, the D -part simplex S^D , to real space, R^{D-1} or R^D , using log-ratio transformations. Descriptions of the properties and processes of these transformations can be found in Pawlowsky-Glahn et al. (2015b).

3.3.1. The centered log-transformed data (clr)

The clr transformation for a composition $x=[x_1, x_2, \dots, x_D] \in S^D$ is the transformed data $y \in R^D$ with:

$$y=\text{clr}(x) = (\ln \frac{x_1}{g(x)}, \ln \frac{x_2}{g(x)}, \dots, \ln \frac{x_D}{g(x)}) \quad (1)$$

That is, each component is represented as a log-ratio of a central value by using the geometric mean of the composition x , $g(x)=[x_1 \cdot x_2 \cdot \dots \cdot x_D]^{1/D}$. The clr-transformed coordinates are always dependent on the geometric mean, so cannot be simply interpreted independently of other elements. Thus, just as in individual components, any single component of the clr-transformation will be associated with the “total”, or, the set of all considered components (Pawlowsky-Glahn et al., 2015a). The main advantages of the clr coefficients are: 1) the clr coefficients translate perturbation and powering of compositions into ordinary sum and multiplication by a scalar of vectors of clr coefficients; and 2) classical Euclidean distance between vectors of clr coefficients is equal to the Aitchison distance of their corresponding compositions (Egozcue and Pawlowsky-Glahn, 2006). These two properties make the clr coefficients extremely useful. The Aitchison distance and the other metric properties including the scale invariance, permutation invariance, and sub-compositional coherence may be effectively related by the use of clr coordinates. Further discussion can be found in Barceló-Vidal and Martín-Fernández (2016). The main disadvantage of the clr transformation is that the clr covariance matrix is singular. Despite this complication, the relationship between the chemical elements can be freely analyzed using the clr-biplot (Aitchison and Greenacre, 2002), and it is for this reason that clr coordinates were used in our analysis and statistical computations.

One of the best methods to display the variable associations and their main behaviour in multivariate analysis is by using a dendrogram arising from hierarchical clustering. Therefore a hierarchical cluster was built based on lognormal data and clr logratio transformation (Fig. 3) to better visualize several

clear variable associations that can be explained from knowledge of their geochemical behaviour, or nearby anthropogenic processes in the survey area.

3.3.2. Clr-biplot

The most common graph to portray the relationship between variables is the biplot (Gabriel, 1971; Kempton, 1984). The biplot is a 2D graphical representation based on the singular value decomposition (Eckart and Young, 1936), where the individuals or sample observations are expressed as dots and the variables as rays. It permits to relate the correlation and the association between variables, and the construction of a smooth distribution of elements depending of their characteristics. When the biplot is used with clr log-transformed data, it facilitates the analysis of all geochemical variables (Aitchison and Greenacre, 2002; Otero et al., 2005). In addition, the lengths of the rays and the angles between rays and between links may be interpreted in statistical terms (Pawlowsky-Glahn et al., 2015b). The coordinates of the samples represented in the clr-biplot are the coordinates of the centered data. Consequently, the Euclidean distance between two samples in the clr-biplot is an approximation of the Aitchison distance between the corresponding compositions in the simplex. The squared length of a ray is proportional to the clr-variance of the corresponding chemical element. The squared length of the link between two vertices of rays is approximately equal to the variance of the log-ratio of the parts corresponding to the clr-rays. In this sense, the closer the vertices are, the more proportional the concentrations of the chemical elements are (Aitchison and Greenacre, 2002). The cosine of the angle between two links approaches the correlation coefficient between the corresponding simple log-ratios. Consequently, the orthogonality of the links suggests that there is no correlation of the log-ratios; it is not possible to estimate the covariance matrix in order to obtain robust counterparts of loadings and scores for the construction of a robust compositional biplot (Filzmoser and Hron, 2011).

3.3.3. Interpolated factor score maps

The interpolated factor association maps have been obtained using ArcGIS™ 10.2 Spatial Analyst software, using Inverse distance weighted (IDW) as the interpolation method. This method assumes that that value of neighbouring observations contribute more to interpolated values of distant observation even if it ignores the variance of values (Cheng, 2008). A distance of 1.5 km was chosen as the linear-weighted combination distance between sample points. Moreover, the structural geology, the environment, and landscape morphologies and characteristics have been taken into account to better visualize spatial patterns of elements associations into the survey area. The different factors score maps obtained were studied and interpreted in accordance with their hypothetical origin (geogenic, anthropogenic or mixed).

3.4. Factor analysis (FA)

Factor analysis is a statistical method that characterizes different groups of chemical elements with approximately similar geochemical patterns (Miesch Programs, 1990). Multivariate statistical analysis was applied to the lognormal of raw data (NRD) and the clr log-transformed data in order to make in evidence the suitable interrelationships between variables clr log-transformed. One objective of this study is to determine the relationships between the 26 variables in the 319 samples of this study using factor analysis. To determine the distribution and the correlation structure of the variables, we employed multivariate statistical analysis by using the principal factor analysis, which clearly identifies unique factors that have a completely different behaviour to the majority of all other factors. This method aids another objective of this study, the determination of the geogenic or/and anthropogenic source of the variables (Reimann and De Caritat, 2005). To better visualize the correlation between the variables, the number of elements considered was reduced based on 3 criteria: 1) remove elements which concentrations are lower than the limit of detection (LOD); 2) choose arbitrary mostly two representative elements (e.g., Rare Earth Elements; REE), which are geochemically congruent; 3) choose elements with a communality of extraction higher than 0.5 (50%) and/or a common variances <0.5 (Table 2).

The factor association maps were carried out using lognormal data and the clr log-transformation.

Details of the 26 elements included in the final variables matrix (Ag, As, Au, B, Ba, Ca, Cd, Co, Cr, Fe, Hg, K, La, Li, Mg, Na, Ni, P, Pb, Rb, Th, Ti, U, V, Zn, Zr) are shown in Table 2. Based on the multivariate analysis, four factor associations (F1, F2, F3, and F4) were derived for both the lognormal data and the clr-transformed data (Table 3).

Table 2. Statistical parameters and descriptive statistics of the 26 elements included in the final variables matrix after varimax orthogonal rotation from soil sediment samples (n=319) of the Sarno River Basin (values presented in mg/kg).

Variable	Unit	samples	Minimum	Maximum	Mean	Median	Std Deviation	Skewness	Kurtosis
Ag	mg/kg	319	0.03	0.63	0.14	0.12	0.09	2.55	9.03
As	mg/kg	319	5.82	111.50	14.65	13.10	8.21	8.68	93.18
Au	mg/kg	319	0.0003	0.21	0.01	0.01	0.02	7.98	91.35
B	mg/kg	319	05.00	46.00	21.86	22.00	7.43	0.19	-0.34
Ba	mg/kg	319	294.50	1034.90	645.50	643.00	155.25	0.08	-0.51
Ca	mg/kg	319	4300.00	127700.00	38739.00	31500.00	23874.00	1.39	1.85
Cd	mg/kg	319	0.15	11.06	0.62	0.54	0.82	11.26	134.48
Co	mg/kg	319	06.00	16.60	12.14	12.30	2.19	-0.31	-0.33
Cr	mg/kg	319	5.30	808.40	35.96	16.80	76.65	5.99	43.19
Hg	mg/kg	319	0.01	0.61	0.09	0.06	0.07	2.77	12.36
K	mg/kg	319	4000.00	47600.00	23907.00	23300.00	10705.00	0.02	-1.16
Fe	mg/kg	319	15700.00	41200.00	29865.52	30100.00	4222.02	-0.45	0.25
La	mg/kg	319	19.70	86.70	46.00	44.70	11.37	0.63	0.55
Li	mg/kg	319	8.40	45.50	21.09	19.80	7.10	0.85	0.42
Mg	mg/kg	319	3300.00	46900.00	10317.00	8700	5557.00	2.83	10.48
Na	mg/kg	319	460.00	15360.00	6444.00	6640.00	3583.00	-0.04	-1.28
Ni	mg/kg	319	8.50	29.40	15.94	15.60	2.91	1.32	3.40
P	mg/kg	319	460.00	6170.00	2720.00	2680	1129.00	0.32	-0.27
Pb	mg/kg	319	25.41	585.16	71.72	63.47	41.13	6.90	76.36
Rb	mg/kg	319	51.20	259.40	171.31	174.20	49.42	-0.27	-0.91
Th	mg/kg	319	3.60	47.80	14.14	13.20	5.27	1.91	7.62
Ti	mg/kg	319	620	2670.00	1632.00	1620.00	342.00	0.17	0.26
U	mg/kg	319	1.80	23.70	4.98	4.60	2.16	3.16	19.90
V	mg/kg	319	43.00	131.00	91.06	90.00	17.12	-0.06	-0.32
Zn	mg/kg	319	60.50	1115.50	160.47	140.60	93.58	4.22	34.42
Zr	mg/kg	319	7.40	269.80	37.59	28.80	29.12	4.03	23.64

These factor associations were considered and interpreted in accordance with the presumed origin, or reason for the variable association, (i.e., geogenic or anthropogenic sources).

Table 3. Varimax-rotated factor (four-factor model) of 26 variables for 319 soil samples from the Sarno basin; table shows the rotated component matrix of the lognormal data (left table) and the clr transformed data (right table). Bold entries: loading values over |0.50|.

variables	Factors				Communalities	
	F1	F2	F3	F4	Initial	Extraction
Ag	0.10	-0.18	0.73	0.32	1.00	0.68
As	-0.22	0.36	0.12	0.64	1.00	0.60
Au	0.04	-0.07	0.62	0.05	1.00	0.59
B	0.77	-0.22	0.22	-0.14	1.00	0.70
Ba	0.89	0.23	0.09	-0.05	1.00	0.86
Ca	-0.22	-0.68	0.24	-0.20	1.00	0.61
Cd	-0.03	0.03	0.56	-0.03	1.00	0.52
Co	0.79	0.40	-0.11	0.27	1.00	0.87
Cr	-0.16	-0.10	0.10	0.83	1.00	0.73
Fe	0.44	0.77	-0.17	0.26	1.00	0.88
Hg	-0.06	-0.19	0.59	0.19	1.00	0.51
K	0.95	-0.11	0.00	-0.14	1.00	0.94
La	-0.33	0.85	-0.08	0.15	1.00	0.87
Li	-0.58	0.68	-0.13	0.23	1.00	0.87
Mg	-0.43	-0.52	-0.01	-0.06	1.00	0.56
Na	0.87	-0.17	0.02	-0.20	1.00	0.83
Ni	0.05	0.13	0.12	0.79	1.00	0.65
P	0.73	-0.19	0.39	0.01	1.00	0.72
Pb	-0.01	-0.04	0.70	-0.06	1.00	0.50
Rb	0.93	0.01	-0.03	-0.11	1.00	0.87
Th	-0.28	0.87	-0.15	0.00	1.00	0.86
Ti	0.19	0.79	-0.26	0.03	1.00	0.72
U	0.32	0.65	0.16	-0.17	1.00	0.58
V	0.81	0.46	-0.16	0.09	1.00	0.89
Zn	0.31	-0.14	0.64	0.04	1.00	0.52
Zr	-0.51	0.73	-0.14	-0.12	1.00	0.83
Eigenvalues	7.41	6.15	2.96	1.55	-	-
% Variance	28.51	23.67	11.39	5.98	-	-
Cumulative %	28.51	52.18	63.57	69.54	-	-

Clr (variables)	Factors				Communalities	
	F1	F2	F3	F4	Initial	Extraction
Ag	0.22	-0.71	-0.08	0.22	1.00	0.61
As	-0.74	0.23	0.26	0.10	1.00	0.67
Au	0.28	-0.75	-0.03	0.03	1.00	0.65
B	0.82	-0.06	-0.08	-0.17	1.00	0.72
Ba	0.72	0.36	0.36	0.07	1.00	0.79
Ca	0.25	-0.18	-0.86	0.01	1.00	0.83
Cd	-0.07	-0.64	-0.04	0.17	1.00	0.54
Co	0.39	0.67	0.30	0.46	1.00	0.90
Cr	-0.18	-0.27	-0.16	0.72	1.00	0.65
Fe	-0.14	0.81	0.27	0.37	1.00	0.88
Hg	-0.15	-0.71	-0.03	-0.08	1.00	0.54
K	0.95	0.11	0.11	-0.07	1.00	0.94
La	-0.74	0.52	0.22	0.04	1.00	0.87
Li	-0.84	0.39	0.01	0.13	1.00	0.86
Mg	-0.10	0.07	-0.87	0.14	1.00	0.79
Na	0.92	0.08	0.04	-0.12	1.00	0.87
Ni	-0.20	0.19	-0.20	0.84	1.00	0.83
P	0.79	-0.38	0.06	0.05	1.00	0.77
Pb	0.02	-0.69	0.06	0.02	1.00	0.59
Rb	0.84	0.32	0.17	-0.07	1.00	0.84
Th	-0.58	0.63	0.29	-0.01	1.00	0.82
Ti	-0.24	0.79	0.22	0.13	1.00	0.75
U	0.09	0.24	0.72	-0.16	1.00	0.60
V	0.41	0.74	0.32	0.24	1.00	0.88
Zn	0.39	-0.68	-0.07	0.13	1.00	0.63
Zr	-0.76	0.56	0.12	-0.06	1.00	0.90
Eigenvalues	9.30	6.34	2.24	1.62	-	-
% Variance	35.76	24.40	8.62	6.22	-	-
Cumulative %	35.76	60.16	68.77	74.99	-	-

4. Results and discussion

4.1. Hierarchical clustering of lognormal versus clr-transformed data

Figure 3 presents a hierarchical clustering of 26 variables lognormal data versus clr log-transformed data, based on the average linkage between elements method. No clear interrelationships and associations between groups of variables are visible in the hierarchical clustering on log-transformed data (Fig. 3, left plot).

The Ca–Mg and Ni–Cr groups are correlated into the same cluster, and in fact, it seems that possible associations of variables can be simplified or/and reduced using the lognormal data. Performing hierarchical clustering on clr log-transformed data (Fig. 3, right plot) highlights the correlation between variables into a Euclidean geometry which clearly identifies the association of groups of elements in different variables. The variables cluster smoothly based on their most likely geochemical origins or interactions, such as the association of Ca–Mg arising from rudist limestone and carbonate sediments between Solofra and Montoro, and the association between Na–K–P–B and U–Ba–Co–V resulting from pyroclastic rocks of the Mt. Somma-Vesuvius. In addition to the background geology, human activities affecting the survey area can be identified such as through the element association of Ni–Cr and Pb–

Zn–Cd–Hg relating to industrial activities such as the famous tanneries of the Solofrana district (source of Cr and Ni), metal processing operations, and the combustion of petroleum products. It is clear that the cluster analysis was able to distinguish particular groups and associations of variables, allowing us to draw conclusions as to their main geogenic and anthropogenic origins in survey area.

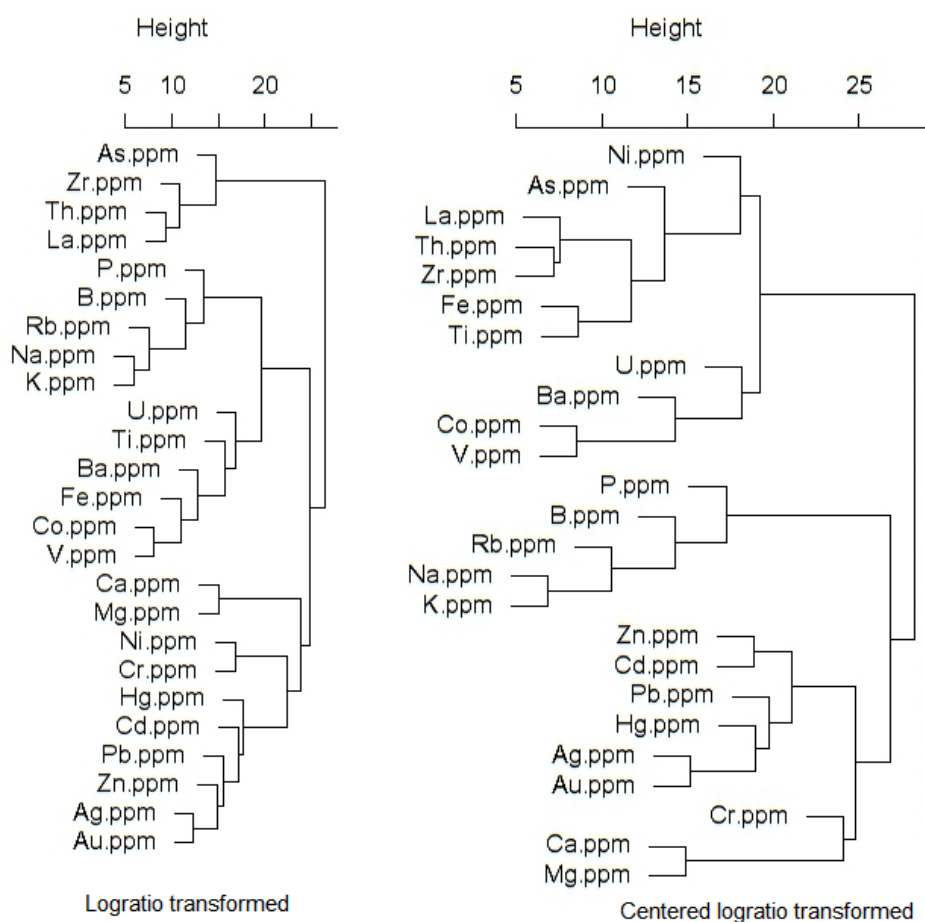


Fig. 3. Dendrograms for lognormal data (left plot) and clr-transformed data (right plot) of the selected 26 variables.

4.2. Generation of clr-biplot

The clr biplot (Fig. 4) explains almost all Euclidean variability (~70%) using only 26 of the total 52 elements analyzed in the first three Principal Components (PC1, PC2, and PC3, accounting for 35.5%, 22.5%, and 10.2%, respectively). Applying this geostatistical computation to the Sarno database displays associations between elements that may reflect their sources, whether related to the underlying geology, or human activity, or a mixture of the two. Following projection into x, y coordinate space (Fig. 4, left plot) we may distinguish five main associations of variables, based on the length of the vertex, their vicinity to each to another and the position between dots (observations or samples) and rays.

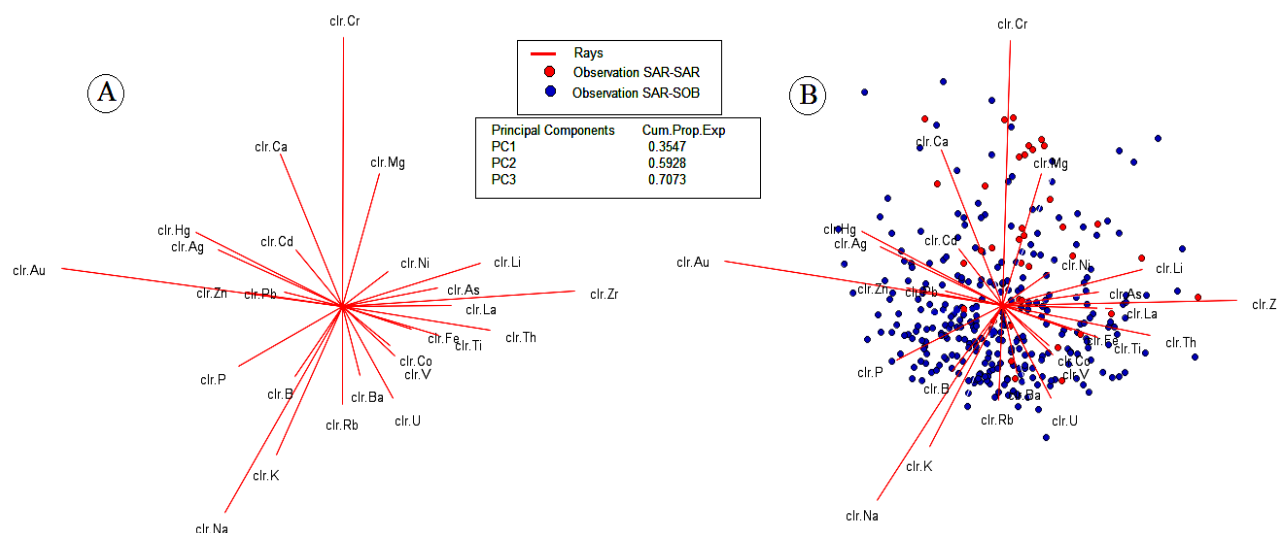


Fig. 4. Clr-biplot of the Sarno Basin soil samples. The left plot (A) is displayed without the individual sample observations, while the right plot (B) shows association between variables and the observations.

The Ca–Mg association (A1) shows both variables with high communalities (the length of the vector) and high correlation coefficient ($r=0.76$). These two elements are commonly found in the carbonate rocks of the study area, such as the silico-clastic deposits in the rudist limestone and carbonate sediment between Solofra and Montoro (see Table 3, F3).

Na–K–P–B–Rb–Ba association (A2) is dominated by a high communality of the Na variable and the vicinity of their rays. Within this group, the correlation coefficient between Na and K is higher ($r=0.91$), followed by Na–P ($r=0.69$), B–P ($r=0.67$) and P–K ($r=0.62$). These associations probably reflect the potassic and ultrapotassic rocks formations that occur throughout the majority of the basin, related to the lava and pyroclastic volcanic activity of Mt. Somma–Vesuvius. Mercury (Hg) presents a high length of communality contributing to the (A3) association with Zn, Pb, Cd, Au, and Ag (see F2 of FA table).

This association may be divided in two sub-groups, named Ag–Au ($r=0.65$) and Zn–Cd–Pb–Hg sub-associations. This association contains both short vectors (Zn, Cd, and Pb are poorly represented) and long vectors (Au, Ag, and Hg). This may, in fact, be representative of the effects of both the geogenic source (volcanic material) and anthropogenic source (the high level of pollution in the basin related automobile gasoline combustion and industrial activities) (Albanese et al., 2013). As–U–Th–Zr–La–Li (A4) and Fe–Ti–Co–V (A5) association vectors display generally the same length of communality, with the exception of Zr (which is significantly greater: 78.8%). Fe is positively correlated with Ti, V, and Co, with correlation coefficients of $r=0.79$, $r=0.78$, and $r=0.65$, respectively. In the A4 association, a similar trend is seen with Th correlating strongly with Zr, La, Li, and U, with correlation coefficients of $r=0.87$, $r=0.85$, $r=0.79$, and $r=0.63$, respectively.

In considering the most likely potential origin of these associations, the A4 association reflects the pyroclastic material covering hilly and mountain areas surrounding the Sarno River Basin, while A5 is most likely related to the predominantly magmatic rocks of Mt. Somma-Vesuvius, with their primary compositions of coexisting Fe–Ti oxides (Cicchella et al., 2014; De Pippo et al., 2006).

Ni and Cr (A6) are modestly correlated ($r=0.58$), and Cr presents a communality value (84.12%). They might reflect a mixing of the main anthropogenic industrial activities of the river basin (i.e., the tannery district of Solofra, metal processing operations, and petroleum combustion).

The blue dots (Fig. 4, right plot) represent the SAR–SOB sampling campaign across the entire basin, while the green dots represent samples collected along the river during the SAR–SAR campaign. It is of note that the SAR–SAR samples are generally correlated to the A6 (Ni–Cr) and A1 (Ca–Mg) vertices. This may suggest that the geochemistry associated with soils in contact with the river, and therefore the river waters themselves, are affected by wastewater from the local industrial activity such as the use of Cr to treat animal skins in the aforementioned tannery industry of the Solofra district.

To clarify these hypotheses, another clr-biplot (Fig. 5) was created based on the amalgamation of the 26 variables in six possible associations (clr-amalgA₁, clr-amalgA₂, clr-amalgA₃, clr-amalgA₄, clr-amalgA₅, and clr-amalgA₆).

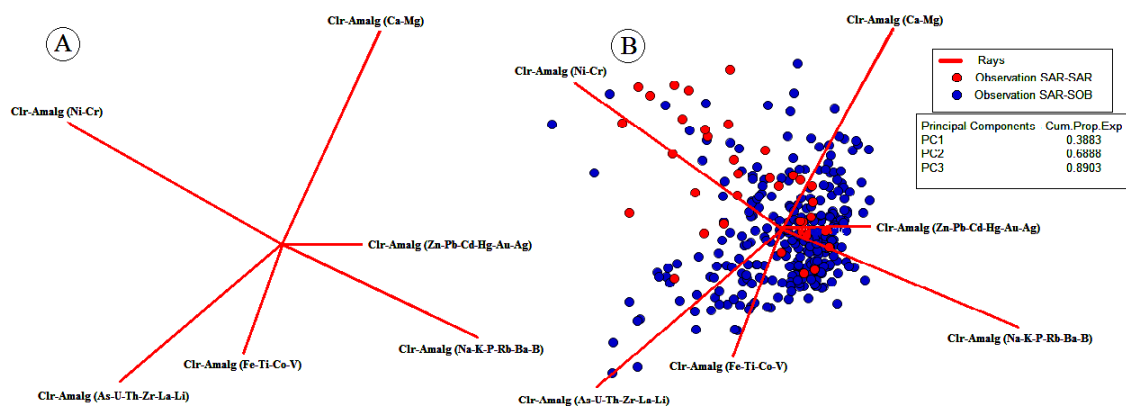


Fig. 5. Clr-biplot of the Sarno Basin soil samples; (A) The clr biplot of amalgamated variables without individual sample observations and (B) with the individual samples included. The amalgamation operation reduced the number of variables based on the initial compositional associations (Fig. 4.). This clarifies the mixed anthropogenic and geological origin of the variables in the survey area.

The amalgamation operation is frequently used to reduce the number of individual parts in compositional dataset, but it is a nonlinear operation in the simplex with the geometry based on the Aitchison geometry (Mateu-Figueras and Daunis i Estadella, 2008). If the D parts of a composition are

separated into $C \leq D$ mutually exclusive and exhaustive subsets, and the components of each subset are added together, the resulting C-part composition is termed an amalgamation (Aitchison, 1986). For example, from the 6-part composition ($x_1; x_2; x_3; x_4; x_5; x_6$) we may obtain the following 3-part amalgamation ($Am1=x_1+x_2; Am2=x_3+x_4; Am3=x_5+x_6$). Elements are assembled into groups which may better explain the underlying associations. However, as pointed out in Pawlowsky-Glahn et al. (2015b, page 52) “amalgamation of parts is problematic when the amalgamated composition is compared with the original one”. In this work, we use the amalgamation algorithm to clarify the possible behaviour of the groups of elements and their main associations. Based on this assertion, the clr-biplot (Fig. 5) was created comprising the amalgamation of the five proposed associations (A1, A2, A3, A4, and A5) based on the compositional clr-biplot (Fig. 4).

The amalgamation of the 26 elements into five association groups reveals important information. The clr-amalgA₁ (Ca, Mg) and clr-amalgA₂ (Na, P, K, B, Rb) may be independent of each other since their rays form 90° angle. This could be due to the different geological substrates of the survey area, confirming the idea that the A1 and A2 associations represent material derived from different lithologies. The same scenario is displayed between clr-amalg (A6) and clr-amalg (A2) which their vertex form a 180° angle and in opposite directions. Clr-amalg (A4: As–U–Th–Zr–La–Li) and clr-amalg (A5: Fe–Ti–Co–V) share closely spaced rays but have differing communalities. This may indicate that they have the same geolithological sources, but result from different volcanic eruptions of Mt. Somma-Vesuvius. Based on the geological background occurring within the basin, A4 and A5 are most likely related to magmatic rocks (pyroclastic rocks) that occur the study area. Nevertheless, we suggest that the different communalities between the two associations may reveal the influence of sub-surface geological phenomenon (magmatism, crystallization) and/or external conditions (weathering, erosion) in the survey area. The high correlation of the Ti–Fe–Co–V association in the study area might be explained by the co-precipitation and absorption phenomena, due to the presence of Fe and Ti hydroxides in the soils developed from pyroclastic covers and volcano-sedimentary rocks.

The clr-amagA6 (Ni–Cr), and amagA1 (Ca–Mg) form an almost 90° angle; this may reflect a different origin, with one (A1) related to the geological sub-strata, and the other (A6) related to anthropogenic activities. Furthermore, A6 may be independent of the other groupings all together, reinforcing a non-geological origin, being related to activities of the tannery industries in the survey area (Albanese et al., 2013). The biplot (Fig. 5, right plot) confirms this interpretation, where the SAR–SAR observations, samples collected along the Sarno River are clearly associated with the clr-amalgA6 ray (Ni–Cr vertex). Moreover, the correlation coefficient of the two variables Ni and Cr is $r=0.58$ (positively correlated). This information indicates that the Ni–Cr contamination is associated with the river and/or is related to a matrix that discharged into the river. It can be confidently attributed to contamination from the tannery

industries (Solofra and Montoro districts), which discharge their wastes directly into the Sarno river and are rich in Cr and enriched in Ni.

4.3. Production of interpolated factor association maps

To show the applicability of clr transformation applied to the factor analysis more clearly, two types of interpolated factor score maps were created, one using the lognormal data (NRD) and the other using the clr-transformed data (CLR). This allows comparison between the two types of data, and will highlight the benefit of using clr-transformed data in showing the distribution of variables in the survey area, and their possible origin. As described above, the different factors obtained from factor analysis were considered and interpreted in accordance with their presumed origin (natural, anthropogenic, or mixed source). A total of 4 factors can be used to account for 69.54% and 74.99% of total cumulative variability in the NRD, and clr-transformed data, respectively (Table 3). This observation confirms that the clr-transformed data more clearly reflects the relationships, associations, and information contained in the original variables. Based on the results of factor analysis, interpolated maps were developed to further clarify the presumed origin of these elements, using frequency space-method (Figs. 6–9). The associations of the four-factor model using lognormal data are:

F1: Na, K, P, Rb, Ba, V, Co, B, $-(Zr, Li)$;

F2: Th, La, Fe, Ti, Zr, Li, U, $-(Ca, Mg)$;

F3: Pb, Zn, Cd, Hg, Au, Ag;

F4: As, Ni, Cr.

For comparison, the associations of the others four-factor model using clr-transformed data are:

F1: Na, K, P, Rb, Ba, B, $-(Zr, Li, As, Th, La)$;

F2: Th, La, Ti–Zr, Fe, Co, V, $-(Pb, Zn, Cd, Hg, Au, Ag)$;

F3: U, $-(Ca, Mg)$; and F4: Ni, Cr.

Elements with loadings over $|0.5|$ are considered representative members of each association factor that results from the chosen factor model, permitting the construction of these interpolated factor association maps.

Fig. 6 and Table 3 display factor scores of variables in the F1 factors of the NRD and clr-transformed data range from -2.79 to 1.80 , and -3.5 to 2.1 , with a percentage of variability of 28.51% and 35.76%, respectively. The factor map (Fig. 6) shows (particularly for the elements Na, K, and P) elevated factor scores near the slopes of Mt Somma–Vesuvius, which decreases gradually towards the inner basin. This may be clearly attributed to the occurrence of igneous units and volcanic soils. These units are dominated by potassic and ultrapotassic lavas and pyroclastic material, which occupies the north-west

Sarno Basin. We can therefore assume, that these maps reflect the concentrations of elements in lithological background of the Sarno Basin due to the pyroclastics rocks of the volcanic soils.

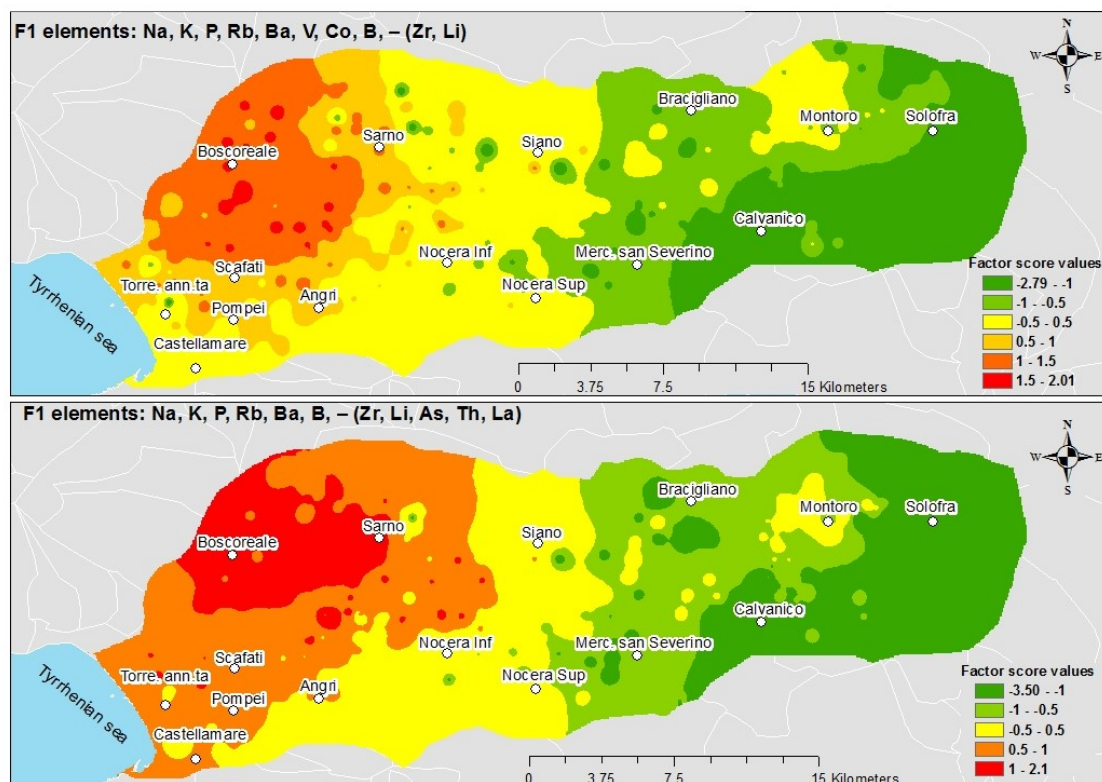


Fig. 6. Interpolated factor score maps showing the elemental association in soils of the survey area based on factor analysis results; (upper) factor scores map F1 (Na, K, P, Rb, Ba, V, Co, B, -(Zr, Li)) using lognormal data, and (lower) factor scores map F1 (Na, K, P, Rb, Ba, B, -(Zr, Li, As, Th, La)) using clr transformed data.

Fig. 7 and Table 3 display factor scores of variables in the F2 factor of the NRD and clr-transformed data range from -2 to 5.9 and -3.57 to 2.62, with a percentage variability of 23.66% and 24.40%, respectively. The construction of these two interpolated factor maps clarifies the behaviour of these elements and their correlations in the basin. In fact, Fig. 7 (upper map) identifies the areas containing the highest factor scores (ranging from 3.95 to 6.0 NDR F2) in the Solofra and Montoro areas and the lowest factor scores are mapped at the slope of the Mts. Picentini and Mts. Lattari. This supports the notion of a lithological background origin to these elements in this area, which is dominated by silicoclastic deposits of the Solofra and Montoro areas, and limestone and dolostone of the Picentini-Taburno unit and Lattari Mountains for Ca and Mg association. The factor score map (Fig. 7, lower) shows the lowest factor scores ranging from -3.57 to -1 and the Pb, Zn, Cd, Hg, Au, and Ag elemental association correspond to the locations of the most urbanized areas (Castellammare di Stabia, Scafati,

Montoro, Calvanico, Nocera, and Solofra) in the Sarno basin. The anthropogenic sources of these variables include the high rates of fossil fuel combustion and vehicle emissions, releasing mostly Pb and Zn into the air and soil (De Pippo et al., 2006; Rossini and Fernández, 2007).

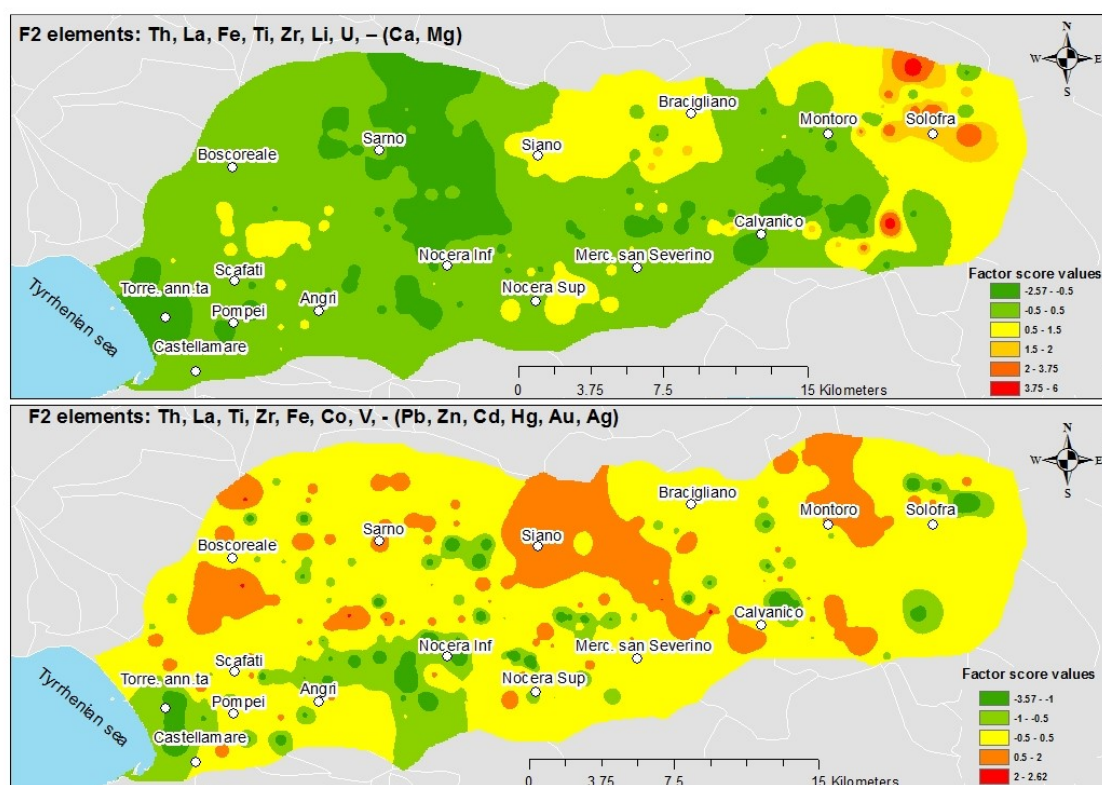


Fig. 7. Interpolated factor score map showing the elemental association in soils of the survey area based on factor analysis results. Upper figure shows the factor scores map F2 (Th, La, Fe, Ti, Zr, Li, U, – (Ca, Mg)) using the lognormal data, and the lower one displays factor scores map F2 (Th, La, Ti–Zr, Fe, Co, V, – (Pb, Zn, Cd, Hg, Au, Ag)) using clr transformed data.

Fig. 8 and Table 3 display factor scores of variables in the F3 factor of the NRD and clr-transformed data range from –1.52 to 5.73 and –3.19 to 3.58 (Fig. 8), with a percentage variability of 11.39% and 8.62%, respectively. The location of hotspots seen in the upper map of Fig. 8 confirms the likely anthropogenic sources of the Pb, Zn, Cd, Hg, Au, Ag elemental associations, which correspond to the location of the main urban centres, while the lower map of Fig. 8 displays the highest factor scores (ranging from 2.5 to 3.58) coinciding with the silicoclastic deposits surrounding the Solofra and Montoro district. Fine-sized solid particles with a charged substrate, such as clay minerals, play an important role in the sorption of trace elements. In fact, U is a highly soluble element and can be easily dissolved, transported and precipitated within sedimentary deposits by subtle changes in oxidation conditions reflecting adsorption and co-precipitation phenomena.

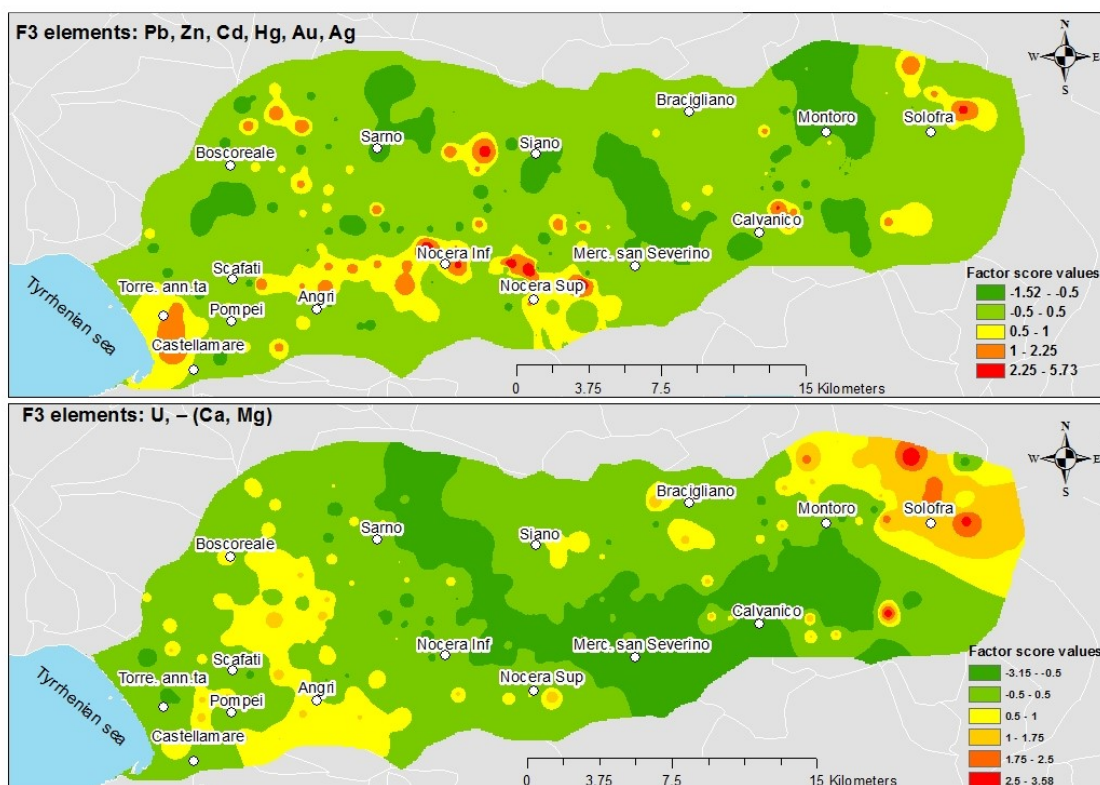


Fig. 8. Factor scores association map in soils of the survey area; (upper) Factor scores map F3 (Pb, Zn, Cd, Hg, Au, Ag) using lognormal data, and (lower) factor scores map F3 (U, - (Ca, Mg)) using clr transformed data.

Fig. 9 and Table 3 display factor scores of variables in the F4 factor of the NRD and clr-transformed data range from 2.48 to 9.85 and 2.09 to 4.05, with a percentage variability of 5.19% and 6.98%, respectively. The highest factor scores values (> 5) for NRD F4 (Fig. 9, upper) are found near the Solofra and Montoro districts, while high factor scores (> 2.01) for CLR F4 (Fig. 9, lower) are located associated with the townships of Solofra, Montoro, Mercato San Severino, Nocera, Scafati, and Castellammare di Stabia (Fig. 9) districts. In fact, all these areas are crossed by the Sarno River or the Solofrana tributaries, which are most likely control the movement of these particles through the study area. The clr transformed data shows a strong correlation of the Cr and Ni elemental all along the Sarno River and its tributaries, which the underlying geology cannot account for. Therefore, it is instead more likely introduced into the environment through anthropogenic activities. Indeed, it is most likely introduced by the release of wastewater from the tannery district of Solofra. This is supported in the literature, where the tannery industrial wastewater is considered to be a major source of pollution, due to presence of chlorophenol and Cr combined with Ni (Tariq et al., 2005). Chromium salts (particularly Cr sulphate) are the most widely used compounds associated with the tanning of skin, while Ni is released from additional tanning agents used during the process and metals industrial activities and the combustion of petroleum products.

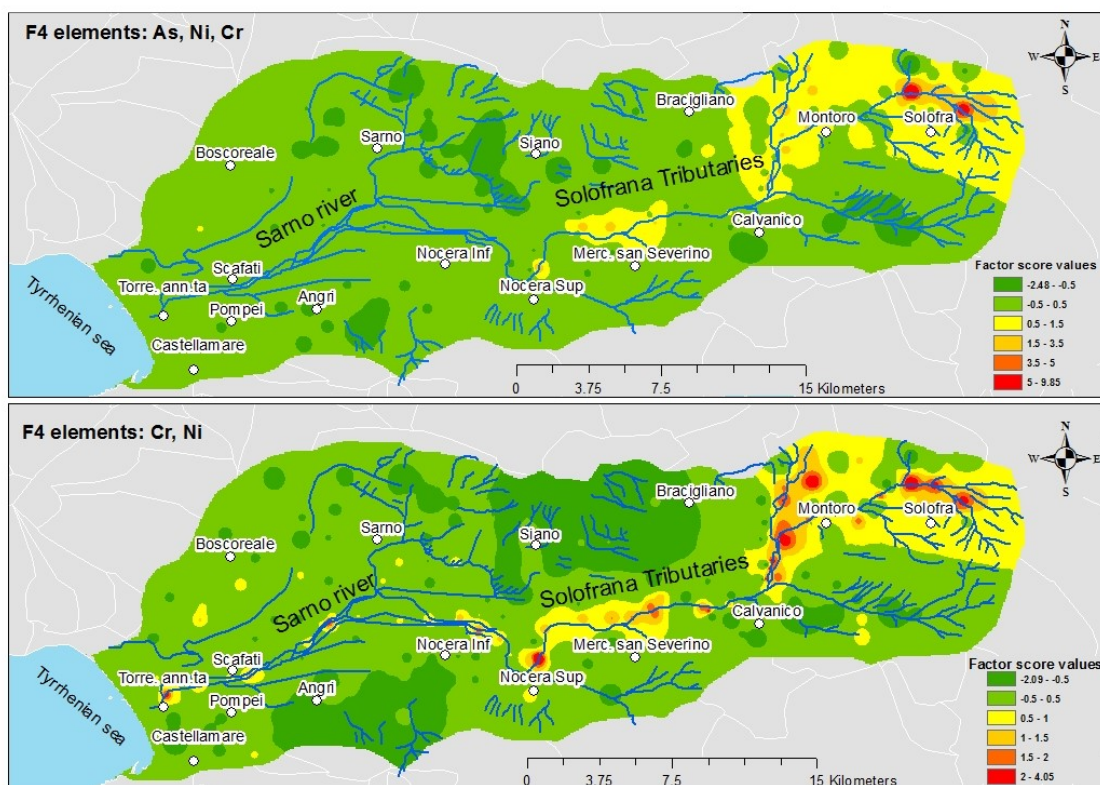


Fig. 9. Interpolated factor score map showing the elemental association in soils of the Sarno River Basin. (Upper) Factor scores map F4 (As, Ni, Cr) using lognormal data, and (lower) factor scores map F4 (Cr, Ni) using clr transformed data.

5. Remarks and conclusion

This study has been the first to use clr-transformed data based on Aitchison distance to interpret the distribution and correlation of multi-elements in the Sarno River Basin, instead of conventional multivariate analysis on raw data (e.g., Albanese et al., 2013; Cicchella et al., 2014).

It shows the applicability of the compositional data analysis, which relates the unique relationships and dependencies between elements that can be lost when univariate and classical multivariate analyses are employed on lognormal transformed data. The combination of use clr-transformation data, hierarchical clustering, principal factor analysis and the construction of interpolated factor score maps allow us to draw the following conclusions:

- It is possible to discern that the anomalies seen in the geochemical distribution of elements in the Sarno Basin have been affected to some degree by both the lithological chemistry (e.g., Pb, Zn, Cd, and Hg in pyroclastic rocks), and the industrial human activities (e.g., Ni and Cr in the tanning and metal industry discharges and emissions) using the clr-transformed coordinates

- The distribution of Ni and Cr (relating to the Ni–Cr association) in the Sarno Basin generally occurs in soil located along the banks of the Sarno River. This cannot be attributed to the underlying geology and is most likely a result of the discharge of wastewater into the river from the tanneries of Solofra. This association is clearer in interpolated factor score maps, using the clr-transformed data.
- The influence of both geogenic and anthropogenic activities on the elevated elemental concentrations in the Sarno River Basin can be clearly identified using the interpolated factor association maps, and results of the clr-biplot using clr-transformed coordinates and after amalgamating variables.
- In addition, the correlation between variables highlighted by factor analysis methods is a useful tool in the interpretation and determinations of the spatial distribution of the patterns, taking into account both their geogenic and anthropogenic sources. Factor association score map NRD F2 in conjunction with CLR F3 (the Ca–Mg association), and factor score map NRD F1 and CLR F1 (the Na, K, P, Rb, Ba, B, Zr, Li, As, Th, and La associations), clearly show high factor scores reflecting the dominant underlying lithology occurring throughout the Sarno River Basin.

Future work could be undertaken to further distinguish the proportionality of the mixed behaviour seen in certain elements (i.e., Pb, Zn, Cd, and Hg), by applying other multivariate statistical tools (such as robust principal component and discrimination analysis) and/or advanced environmental geochemistry methods such as the use of biomarkers. This may continue to highlight the environmental conditions in the Sarno Basin, and provide a framework in which to work to improve the quality of life for resident population.

Section 2.5

Soil geochemical follow-up in the Cilento World Heritage Park
(Campania, Italy) through exploratory compositional data analysis and
C-A fractal model

This section has been published in:
Journal of Geochemical Exploration

Part of special issue:
Multifractals and singularity analysis in mineral exploration and environmental assessments
Volume 189, June 2018, Pages 85-99

Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model

Abstract

Campania Region (southern Italy) is characterized by several elemental soil anomalies both geogenic and anthropogenic. Parts of these anomalies occur into the World Heritage Territory known as “National Park of the Cilento and Vallo di Diano”, where this follow-up study has been carried out. In this paper, a methodology based on compositional data analysis (CoDA) and factor score C-A fractal model was applied on geochemical data from the above study in order to identify geochemical signatures associated anomalies. Eighty-one top soil samples were collected over an area of 98 km², and analyzed by ICP-MS after aqua regia digestion. Frequency based method (edaplots, classical and robust compositional biplot) and frequency space-method (factor score maps) were applied to visualize the correlation between variables and their main features into the survey areas. The different geochemical patterns were distinguished by a multivariate analysis combined with Concentration-Area (C-A) fractal method. Results show that geochemical data should be transformed under a compositional perspective to avoid artefacts, prior to statistical computations. Indeed, ilr-transformed data show a distinct bimodal distribution for several elements. This type of distribution appears masked considering raw and lognormal data. A “robustification” of the variables dataset permits to found a more clear relationship between variables. Factor score maps based on ilr-transformed variables and C-A plot cut off threshold displayed different geochemical patterns. In our survey area soil alteration phenomena could mask the nature of parental rock. The factor scores maps highlight an antithetic behaviour of many elements. Elements characterized by elevated geochemical mobility, such as Ca and Mg, are negatively correlated with Chemical Alteration Index values. Instead, Sn, Th, Be, Al are characterized by elevated geochemical stability; for this reason they are found where high-weathered soils occurs. The presence of elements such as Co, Cu, Fe, Ni, Cr, Zn, K, and Mn is mainly controlled by terrigenous flysch deposition. As, Pb (exceeding the CSC) and Sb association mainly occur in correspondence of urban areas and where traffic jams are frequent.

1. Introduction

In the last two decades, geochemical mapping has been recognized as a relevant method both in mineral exploration and environmental geochemistry. The compositional nature of geochemical data has long been neglected. These types of data are compositional because they belong to a closed system (each element is a part of a whole) and they should be treated as such to avoid spurious correlation

(Pawlowsky-Glahn and Buccianti, 2011a). To treat the compositional geochemical data in univariate or multivariate analysis, we must move the D-part simplex S^D , to real space, $R^D - 1$ or R^D , using log-ratio transformations (additive log-ratio (alr), centered log-ratio (clr) and isometric log-ratio (ilr)) that assume the sample in Aitchison geometry which can properly reconcile the correlation between elements (Aitchison, 1986; Egozcue et al., 2003; Filzmoser and Hron, 2008; Filzmoser et al., 2009a; Filzmoser et al., 2009b; Carranza, 2011; Zuo et al., 2013). Different methods have been used for processing geochemical data and generating spatial distribution interpolated maps from point data. “They belong to frequency-based and frequency-space-based methods” (Zuo et al., 2013). The frequency-based methods are based on the analysis of data frequency distributions using univariate and multivariate tools (Filzmoser et al., 2009a). Among the frequency-space-based methods, fractal and multifractal analysis (Cheng et al., 1994; Mandelbrot, 1983) have been demonstrated to be a powerful tool for identifying geochemical anomalies (e.g., Cheng, 1999a, 2007; Cheng et al., 2000c, 2010; Lima et al., 2003a; Cicchella et al., 2015; Zuo, 2011; Zuo et al., 2015) and separate them from background values. The anomaly analysis is one of the most discussed issues because it has been helpful to improve the quality of geochemical interpretation of natural and anthropogenic phenomena (Cicchella et al., 2005). These exploratory methods of geochemical data have been widespread applied to soil geochemical data analysis. In this study we propose an exploratory compositional data analysis and concentration-area (C-A) fractal model applied on factor scores of the element associations obtained by Factor Analysis to detect the origin of multi-element soil anomalies (geogenic, anthropogenic or mixed). The latter procedure was applied on data from the soil follow-up geochemical survey in the Cilento - Vallo di Diano area, stemming from a regional survey study covering the entire Campania region, southern Italy (Buccianti et al., 2015; De Vivo et al., 2016). Campania region is characterized by several elemental soil anomalies of both geogenic and anthropogenic sources (Albanese et al., 2007; De Vivo et al., 2016; Lima et al., 2003a; Zuzolo et al., 2016). Parts of these anomalies occur into the World Heritage Territory known as “National Park of the Cilento and Vallo di Diano”. Previous regional scale studies on stream sediments (Albanese et al., 2007) and in soils (De Vivo et al., 2016) showed interesting different potentially toxic elements anomalies (PTEs). In particular, for soils in different sites occur values exceeding the contamination thresholds (CSC) established by Italian legislation for soils (Legislative Decree 152/2006, n.d.). The attention on the Cilento - Vallo di Diano area was also individuated by Minolfi et al. (2018a) study, which indicated the presence of several PTEs exceeding the CSC. A geochemical follow-up survey was carried out on the aforementioned area, in order to assess the distribution pattern of geochemical anomalies and to distinguish the natural sources (geogenic) from the anthropogenic ones. The main objectives of this study were: (1) to illustrate the importance of compositional log-transformations in geochemical data for anomaly features assessment; (2) to

delineate their main sources by multivariate analysis and GIS-based approach; (3) to use robust compositional data evaluation to assess geogenic or anthropogenic conditions of multi-element associations when the classical biplot is sensitive to outliers; (4) to prove the usefulness of factor score maps using C-A fractal model for processing geochemical data and to distinguish different geochemical distribution patterns. The geochemical data analysis was performed taking account both of the compositional and their fractal distribution nature.

2. Univariate analysis and compositional biplot

To better visualize the element distributions and their main different sources into the survey area, univariate and multivariate computations were displayed through edaplots and compositional biplot, respectively. Edaplot (Exploratory Data Analysis Plot) is a combination of histogram, density plot, on dimensional scattergram and a boxplot and is considered one of the best statistical tools to display the data distribution (Reimann et al., 2008). The histogram is able to display unimodal (symmetric distribution or skewed) or multimodal distribution.

The one-dimensional scattergram itself draws up the data along a straight line while the density plot traces a variation of the histogram that uses kernel smoothing to plot values; the picks and the form of the shape allow the interpretation of the data distribution.

Compositional biplot is a powerful statistical tool which displays both samples (observations) and the variables of a data matrix in terms of the resulting scores and loading (Gabriel, 1971). Thus the scores represent the structure of the compositional data hold into a Euclidian space based on variance and covariance matrix; moreover they display the association structure of the dataset where the rays are define from the center of the plot and the length of vectors are proportional to the amount of explained variance (communality) of the variables they represent.

The interpretation of the graphic depends on the loading (rays) structures and in more details the approximate links between rays and samples, the distances between vertexes and their directions (Otero et al., 2005). The utility of compositional biplots is better described in several papers (Pison et al., 2003; Maronna et al., 2006; Filzmoser and Hron, 2008; Filzmoser et al., 2009b; Otero et al., 2005; Hron et al., 2010) related to their use on compositional data elaboration.

Classical compositional biplot (CCB) and Robust Compositional Biplot (RCB) are used to compute the correlation between variables using compositional raw data and log-transformed data. However the biplot of raw data is substantially influenced by occurrence of outliers and the fact that it misleads the compositional nature of the data matrix which can affect the principal components in results interpretation (Aitchison, 1986; Filzmoser et al., 2009a, 2009b). For these reasons log-transformed data

are recommended to use in multivariate analysis, in addition to the robust version of compositional biplot..

3. Concentration-Area fractal model (C-A plot in factor scores maps)

Factor analysis was used to investigate the complex multivariate relationships among variables, which are not properly put on display by simple correlation analysis. Factor analysis is a multivariate statistical technique for identifying different groups of chemical elements with approximately similar geochemical patterns. It extracts the most important information from the data on the concept of communality (for each variable, communality is defined as the common variance explained by the factors).

Cheng et al. (1994) proposed a concentration-area (C-A) fractal model for geochemical anomaly separation. From a multifractal point of view, the extreme (anomalous) values may follow a fractal distribution rather than a normal or lognormal law. The concentration-Area (C-A Plot) fractal method separates anomalies from background on the basis of concentration values, as well as the spatial and geometrical properties of geochemical patterns. Point data can be interpolated by multifractal-IDW (MIDW) interpolation based on GeoDAS software (GeoData Analysis System for Mineral Exploration and Environmental Assessment) (Cheng, 1999b, 2000a; Lima et al., 2003a). A C-A plot on log-log paper can be used to establish power-law relationships between the area $A(\geq s)$ with the concentration values greater than s and the concentration value s itself. On a log-log paper a fractal curve is calculated and breakpoints (change in slope of the curve) represent different statistical populations. Several straight-line segments can be fitted in correspondence of breakpoints to provide a set of cut-off values for subdividing the concentration scale into discrete classes. These discretizations define different geochemical population interpretable as specific features of the area.

In this case, C-A fractal method was applied on the elemental association factor scores to highlight anomalies of different geochemical processes into the survey area.

4. Materials and methods

4.1. Survey area

The study area covers an area of 98 km² in south-west of Italy and settles in the southern part of the Campania Region. Most of the survey area falls within a World Heritage Territory known as “National Park of the Cilento and Vallo di Diano”. The geomorphology of the area is predominantly mountainous and hilly constituted by Cervati Mountain (the highest mountain of Campania region with 1899 m high). Most of the survey area is drained by Bussento and Mingardo rivers (Fig. 1A) and is mainly characterized by the presence of natural outlines such as grasslands and persistent woodlands (Fig. 1B).

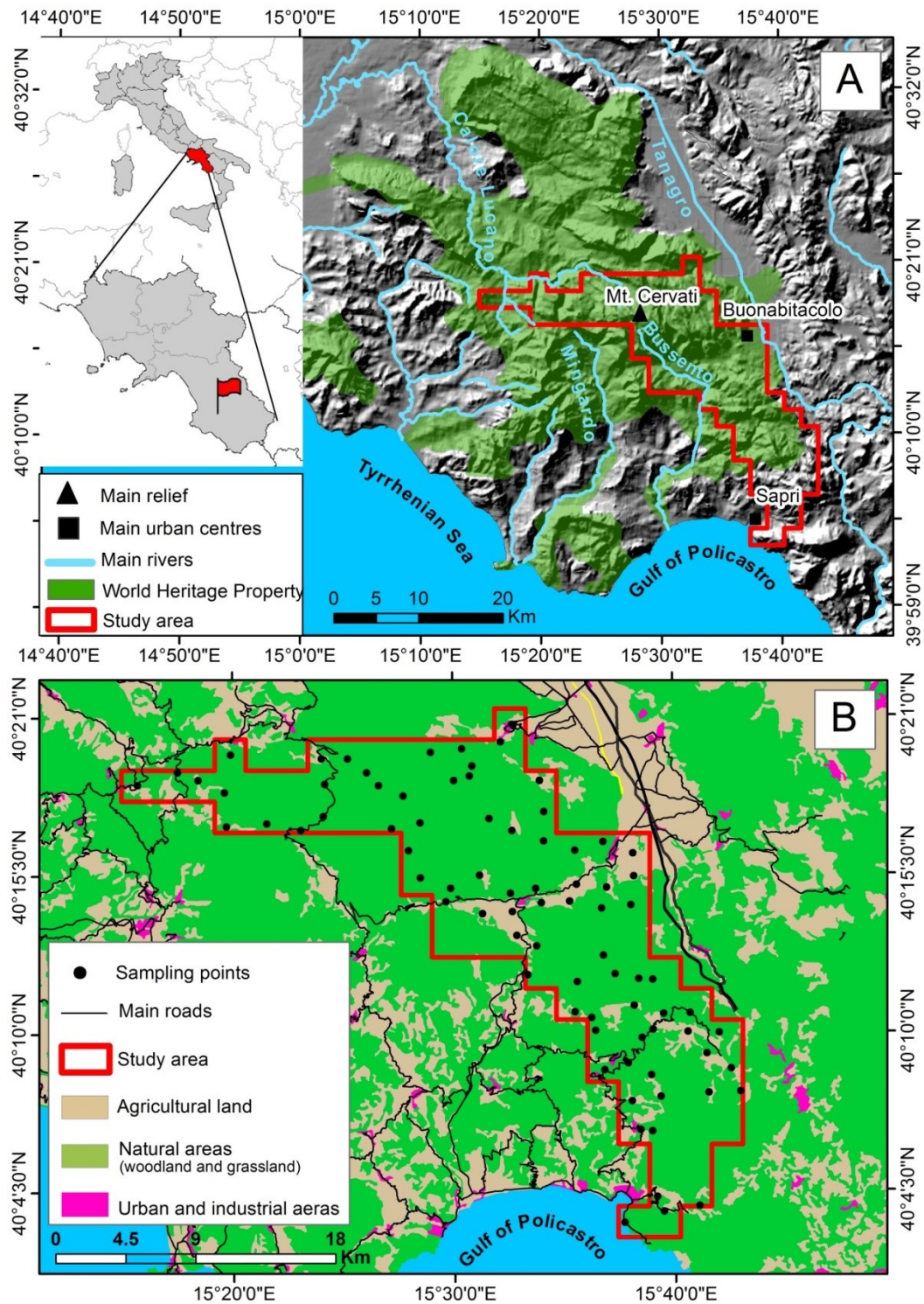


Fig. 1. Introductory map: Localization and main geomorphological features of the study area. B Land use, road map and sampling point location of survey area (Cilento, Southern Italy).

Heterogeneous agricultural areas are mainly concentrated in less steep areas (Cilento plain) with extensive orchards, vineyards, and chestnuts, greenhouse horticulture, floriculture, and crop production throughout and around the main urban zones (Buonabitacolo, Sapri). The survey area is crossed by

different trunk roads and the A3 highway which skirts the area and connects the main urban centers. Urban environments are limited, discontinuous and concentrated mostly along the coastal side. The survey area is part of the Southern Apennines fold and thrust belt formed as a result of the convergence between the Apulian and the European plates since the late Cretaceous (Vitale et al., 2010). In the soils of Italy, as well as in our survey area, the nature of the parental rock can often be masked by the effects of climate and vegetation and alteration degree, that play a fundamental role in determining soil properties and development (Costantini and Dazzi, 2013). For this reason and in line with the purposes of this study, we propose a land system map in Fig. 2 (mod. from Di Gennaro et al., 2002). This map highlights areas with typical “patterns” of lithology and morphology; the occurring patterns are governed by climate, geology and landform which interact over time to influence the distribution of soils and vegetation.

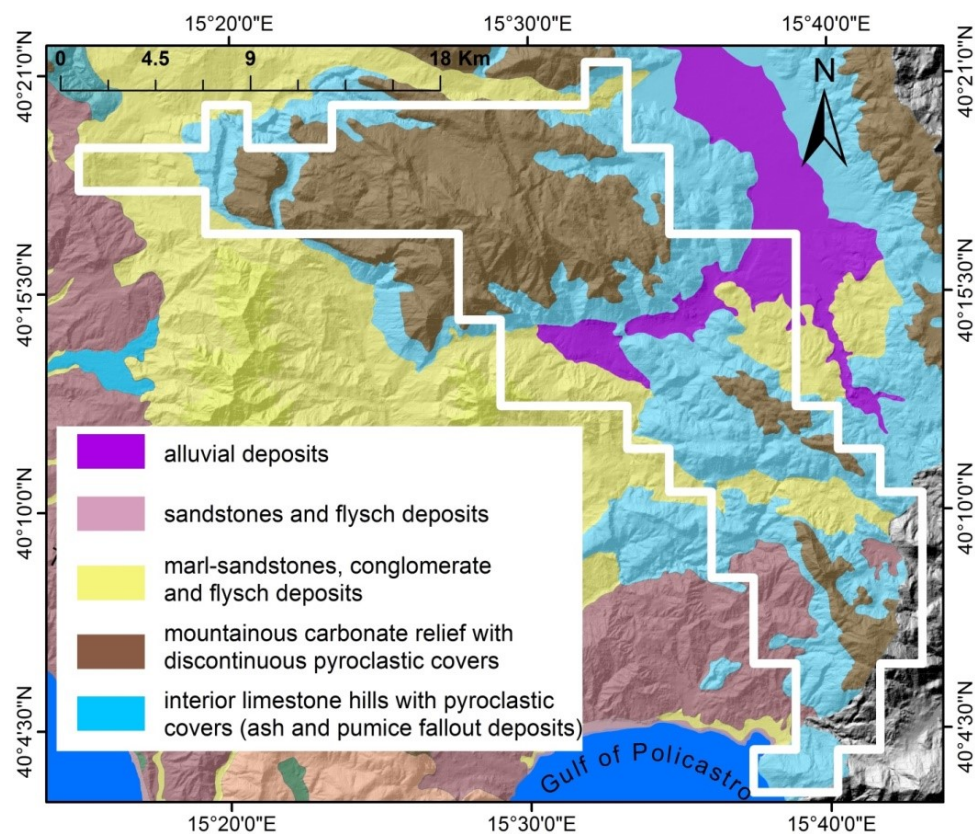


Fig. 2. Land system map of the study area (Cilento, southern Italy). Modified from Di Gennaro et al., 2002.

It is possible to observe the predominance of mountainous carbonate reliefs with discontinuous pyroclastic coverings and widespread limestone hills controlled by ash and pumice fallout deposits.

The Quaternary volcanic activity of Vesuvius (100 km NW), Phlegrean Fields (Campi Flegrei) (120 km NW), Roccamonfina (200 km NW), Mt. Vulture (located east of the Apennine mountain range, about 100 km NE) and Aeolian islands volcanic arc has exerted an influence on the formation of soils in the study

area, giving a widespread pyroclastic footprint (Buccianti et al., 2015; Scheib et al., 2014). Moreover, marl-conglomerate, marl-sandstone and alluvial deposits occur in the study area.

4.2. Sampling and sample preparation

During summer 2016, 81 topsoil samples were collected from the study area (98 km²) at an average sampling density of approximately one sample per 1.2 km². About 1.5 kg soil samples was collected at a depth between 5 and 15 cm under the ground surface after the vegetation cover removal, following the protocol according to the FOREGS sampling procedures (Plant et al., 1996; Salminen et al., 1998). Each sample is made up from a composite soil material taken from five points over 5 m square and in every 20 sampling sites a duplicate sample was collected in the same cell to allow the blind control of the cell sampling variability combined with analytical variability. At each sampling spatial coordinates, topography, local geology, type and main properties of soils, land use, and any additional detail related to anthropic activities in the surroundings were recorded. The 81 samples were dried with infra-red lamps at a temperature below 35 °C, pulverized in a ceramic mortar and then sieved to retain the < 2 mm fraction. The pulps were stored in small plastic bags containing at least 30 g of samples for chemical analysis.

4.3. Chemical analysis

Analyses were carried out by Mineral Laboratories Bureau Veritas (Vancouver, Canada). Each sample was digested in aqua regia solution and analyzed by ICPMS (inductively coupled plasma-mass spectrometry) for 53 elements as required by Italian environmental Law (D. Lgs. 152/2006). The quality of data was assessed by estimation of accuracy and precision (Table 1).

Precision of the analysis was calculated using in-house replicates, while accuracy was determined using laboratory's in-house reference material.

Table 1. Detection limit, accuracy and precision of the applied analytical method (RPD =relative percent difference; %DL= percentage of samples with concentrations below the detection limit). ^a Precision was calculated as relative percentage difference (%RPD) using the formula: $\%RPD = \frac{|SV - DV|}{SV + DV / 2} \times 100$, where SV =the original sample value, DV =the duplicate sample value. ^b The laboratory accuracy error was determined using the formula: $Accuracy\ error = \frac{|X - TV|}{TV} \times 100$, where X =laboratory's analysis result for the performance sample (standard) and TV= true value of the performance sample (standard).

Elements	Unit	Detection limit (DL)	%<DL	Precision ^a (% RPD)	Accuracy ^b (%)
Al	%	0.01	0	3.1	3.4
Ca	%	0.01	0	3.3	4.6
Fe	%	0.01	0	2.2	3.4
K	%	0.01	0	1.9	2.4
Mg	%	0.01	0	1.9	3.0
Na	%	0	0	1.4	6.0
P	%	0	0	5.9	7.4
S	%	0.02	6	5.2	1.2
Ti	%	0	2	3.4	9.2
As	mg/kg	0.1	0	7.8	4.8
B	mg/kg	1	0	2.0	9.0
Ba	mg/kg	0.5	0	10.0	5.3
Be	mg/kg	0.1	0	21.2	18.0
Bi	mg/kg	0.02	0	6.7	10.1
Cd	mg/kg	0.01	0	7.4	4.5
Co	mg/kg	0.1	0	2.1	1.0
Cr	mg/kg	0.5	0	2.4	5.6
Cu	mg/kg	0.01	0	7.2	4.3
Ga	mg/kg	0.1	0	5.8	4.4
La	mg/kg	0.5	0	5.7	10.1
Li	mg/kg	0.1	0	2.1	2.2
Mn	mg/kg	1	0	1.7	6.0
Mo	mg/kg	0.01	0	3.1	4.0
Ni	mg/kg	0.1	0	3.6	2.1
Pb	mg/kg	0.01	0	3.5	5.0
Sb	mg/kg	0.02	0	7.4	7.3
Se	mg/kg	0.1	2	15.2	7.3
Sn	mg/kg	0.1	0	4.2	12.5
Sr	mg/kg	0.5	0	5.1	10.0
Th	mg/kg	0.1	0	3.5	8.4
Tl	mg/kg	0.02	0	1.1	3.0
U	mg/kg	0.1	0	13.6	11.9
V	mg/kg	2	0	4.5	3.1
W	mg/kg	0.1	14	3.0	2.0
Zn	mg/kg	0.1	0	5.7	8.1
Ag	µg/kg	2	0	4.5	24.9
Au	µg/kg	0.2	2	3.6	17.7
Hg	µg/kg	5	0	7.0	4.2

4.4. Data processing and factor score maps

In this paper, the production of univariate (Edaplots) and multivariate (Compositional biplot) graphics was implemented by the free and open source statistical software R, one the most used statistical tools to provide descriptive statistics and graphics to advanced methods (e.g. robust compositional tools).

Two mains open-source packages for R were used for compositional data analysis: “Compositions” (Van Den Boogaart et al., 2011) and “Robcompositions” (Templ et al., 2011).

The univariate computations (Edaplots) were displayed using the raw data, lognormal data and ilr log transformed data of As, Pb and Co. The ilr log transformation was applied on data taking into account the compositional vectors of n parts partitioned into groups of parts presenting a certain affinity (Egozcue et al., 2003). “It is based on the choice of an orthonormal basis (in the well-known Euclidean sense) in the hyperplane formed by the clr transformation.” (Filzmoser and Hron, 2009). The orthonormal basis were built defining non-overlapping groups through a sequential binary partition (SBP) of the whole composition (Filzmoser and Hron, 2009; Pawlowsky-Glahn and Egozcue, 2011). The coordinates which represent an element of the simplex in the orthonormal basis defined by an SBP represents the balances (Egozcue and Pawlowsky-Glahn, 2005). This computation is shown in Table 2 which contains 18 different elements or parts in the simplex S18, and this information can be expressed within 17 dimensions in R17 that will form the balances.

Table 2. Sequential binary partitions and resulting balances of 18 elements of the survey areas. Elements are grouped referring to the hierarchical cluster which takes into account their affinity.

Balance	Ca	Mg	Zn	P	K	Th	Sn	Al	Be	As	Pb	Sb	Ni	Cr	Fe	Cu	Co	Mn
Z ₁	+	-																
Z ₂			+	-														
Z ₃					+	+	+	-	-									
Z ₄					+	+	-											
Z ₅					+	-												
Z ₆								+	-									
Z ₇										+	+	-						
Z ₈										+	-							
Z ₉													+	+	+	-	-	-
Z ₁₀													+	+	-			
Z ₁₁													+	-				
Z ₁₂																+	-	
Z ₁₃																	+	-
Z ₁₄	+	+	-	-														
Z ₁₅	+	+	+	+	-	-	-	-	-									
Z ₁₆	+	+	+	+	+	+	+	+	+	-	-	-						
Z ₁₇	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-

In this study, the affinity throughout variables or “partition of elements” was assessed by applying a hierarchical cluster analysis (Fig. 3) which displays links between parts.

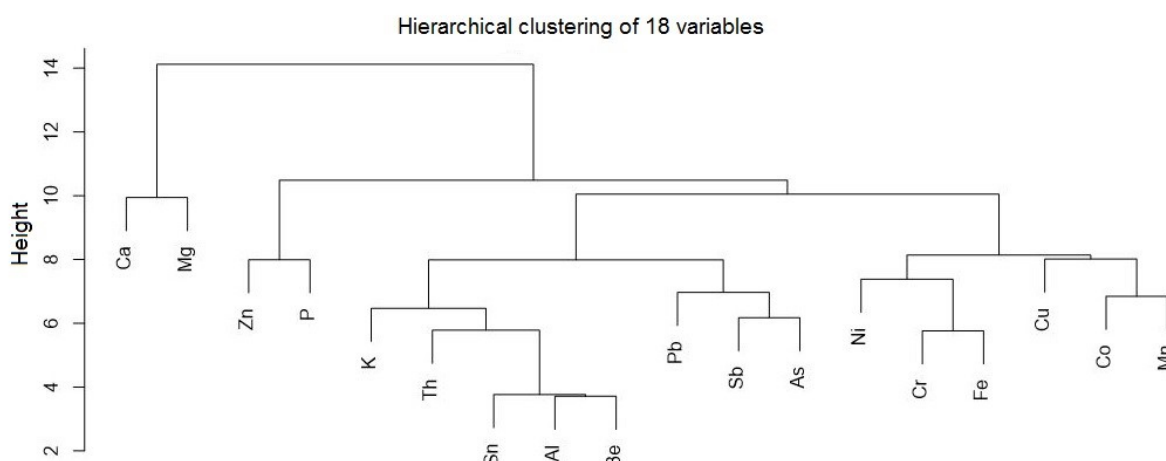


Fig. 3. Hierarchical cluster of 18 variables shown as a dendrogram which displays the relationships between elements and their affinity; this cluster is plotted using the average linkage between variables.

For the multivariate computations, such as robust biplot and factor analysis, it is usefulness to use balances (Z_1 , Z_2 , etc....) loadings and scores in plot because is not straightforward to interpret in sense of composition. Indeed, the variables can be retraced applying a clr back transformation of the balances (Z_1 , Z_2 , etc....) to their corresponding variables (As, Cu, etc....) in the clr space. This back transformation doesn't affect the covariance matrix between variables due to the fact that it exists a linear relationship between variables ilr and clr transformations (Egozcue et al., 2003) and ilr transformation expresses the clr coordinates in a orthonormal basis (Filzmoser et al., 2009a, 2009b). To better visualize the correlation between variables and display the main behaviour of these variables into the survey areas, 18 elements were considered (Table 3).

The reduction of the number of elements is based on 3 main criteria: 1) removal of elements with > 40% of values below the limit of detection (LOD); 2) choose arbitrary mostly two representative elements (e.g., Rare Earth Elements; REE), which are geochemically congruent; 3) choose elements with a communality of extraction higher than 0.5 (50%) and/or a common variances < 0.5 (e.g. Reimann et al., 2002).

Factor score distribution maps were performed by means of GeoDAS (Cheng, 1999b, 2000a, 2000b) using Multifractal Inverse Distance Weighted (MIDW) script as an interpolation method (Cheng, 1999b) and ArcGIS software. Factor scores have been classified using the concentration–area fractal method (C-A).

Table 3. Statistical parameters for 18 elements of 81 topsoil samples of the survey area.

Element	Unit	Number of samples	Min	Max	Mean	Median	RMS	Std Deviation	Skewness	Kurtosis
Al	mg/kg	81	6300.00	69700.00	36541.98	31900.00	40009.77	16394.63	0.42	-0.92
As	mg/kg	81	2.50	65.10	16.09	14.70	19.52	11.12	1.74	4.57
Be	mg/kg	81	0.20	7.00	3.25	3.10	3.72	1.83	0.36	-1.09
Ca	mg/kg	81	2600.00	288700.00	37007.41	16500.00	61818.42	49825.88	2.81	9.02
Co	mg/kg	81	2.90	27.50	14.29	14.20	15.07	4.83	0.48	0.58
Cr	mg/kg	81	11.50	106.70	41.88	40.20	45.29	17.33	1.38	3.13
Cu	mg/kg	81	8.80	151.23	38.34	35.34	42.26	17.90	3.59	19.25
Fe	mg/kg	81	6200.00	59700.00	32785.19	33100.00	34114.33	9488.44	-0.08	0.17
K	mg/kg	81	800.00	5400.00	3439.51	3600.00	3603.19	1080.36	-0.04	-0.83
Mg	mg/kg	81	2000.00	63300.00	7439.51	5300.00	10957.92	8095.61	4.79	27.47
Mn	mg/kg	81	118.00	2239.00	959.63	916.00	1028.33	371.86	0.67	1.22
Ni	mg/kg	81	9.40	81.50	37.43	37.40	39.45	12.53	0.54	1.42
P	mg/kg	81	100.00	2710.00	835.68	790.00	960.17	475.77	1.54	3.29
Pb	mg/kg	81	9.21	165.10	38.37	40.83	43.49	20.59	2.75	15.72
Sb	mg/kg	81	0.17	3.66	1.04	1.01	1.22	0.64	1.33	3.11
Sn	mg/kg	81	0.50	4.90	2.32	2.20	2.54	1.04	0.44	-0.68
Th	mg/kg	81	1.10	23.70	8.79	6.70	10.24	5.28	1.01	0.21
Zn	mg/kg	81	30.80	392.40	103.21	102.50	111.04	41.20	4.20	28.32

This method allows images to be subdivided into components for symbolizing distinct image zones representing specific features on the ground. Considering that the factor scores values present negative and zero values which are not “log-transformable”, was applied a min-max normalization by scaling the original data within a specified range of features (e.g., ranging from 1 to 100). Furthermore, min-max normalization performs a linear transformation on the original data without changing the geometrical structure of the simplex (Han and Kamber, 2001).

Suppose that X and Y are the minimum and the maximum values for initial dataset S . Min-max normalization maps a value for v of S to v' in the range $[\mu, \partial]$ by computing:

$$v' = \frac{(v-X)}{(Y-X)} \times (\partial - \mu) + \mu \quad (1)$$

where μ and ∂ relate the new minimum and maximum of the specified new range, successively.

Specifically, thresholds were determined by a C-A plot, in which the vertical axis represents cumulative pixel areas $A(p)$, with element concentration values greater than p , and the horizontal axis the normalized factor scores values (p). Initial factor scores values will be obtained by computing a back-

transformed of the normalized factor scores data. Breaks between straight-line segments and corresponding values of p were used to reclassify maps of elemental association factor scores maps.

4.5. Chemical alteration degree

Weathering indices are conventionally used to quantify the extent of weathering within a sample soil and help to explain processes occurring during pedogenesis. In our survey area, as previously described, soil alteration phenomena could mask the nature of parental rock. The extent of alteration was quantified using Chemical Index of Alteration (CIA). This index was proposed by Nesbitt and Young (1982) and has been extensively applied as a proxy to examine weathering. The CIA index is calculated as:

$$CIA = \frac{Al_2O_3}{Al_2O_3 + CaO + Na_2O + K_2O} * 100 \quad (2)$$

The calculation of the CIA is a tool to measure the weathering phenomena of the geological feature of a survey area; it's well established certain range of values (percentage) to describe the level of weathering. In consequence, the raw concentration values are likely the more suitable values for this computation because the log transformed data (coordinates) present null and negative values which are not useful rapport of ratio (on denominator) computation. During weathering the proportion of alumina to alkalis would typically increase in the weathered product and that is a good measure of the degree of weathering. The CIA values for the soil samples varied in the range 2.85–88.90 (average=62.20), showing a high variability of the degree of weathering. In general, CIA values under 50 are characteristic of unweathered soils; 100 correspond to the optimum weathered value. Consequently, soils are considered weathered when clay minerals (e.g. montmorillonites) dominate, generally in the range of 75–85 (Nesbitt and Young, 1982; Price and Velbel, 2003).

5. Results and discussions

The chemical characterization of the survey area soils shows several PTEs (As, Be, Cd, Co, Cu, Pb, Sn, Tl, V and Zn) exceeding the contamination thresholds (CSC) established by Italian legislation for soils (Legislative Decree 152/2006, n.d.). Many of these elemental exceeds are very interesting due their spatial distribution (Fig. 4).

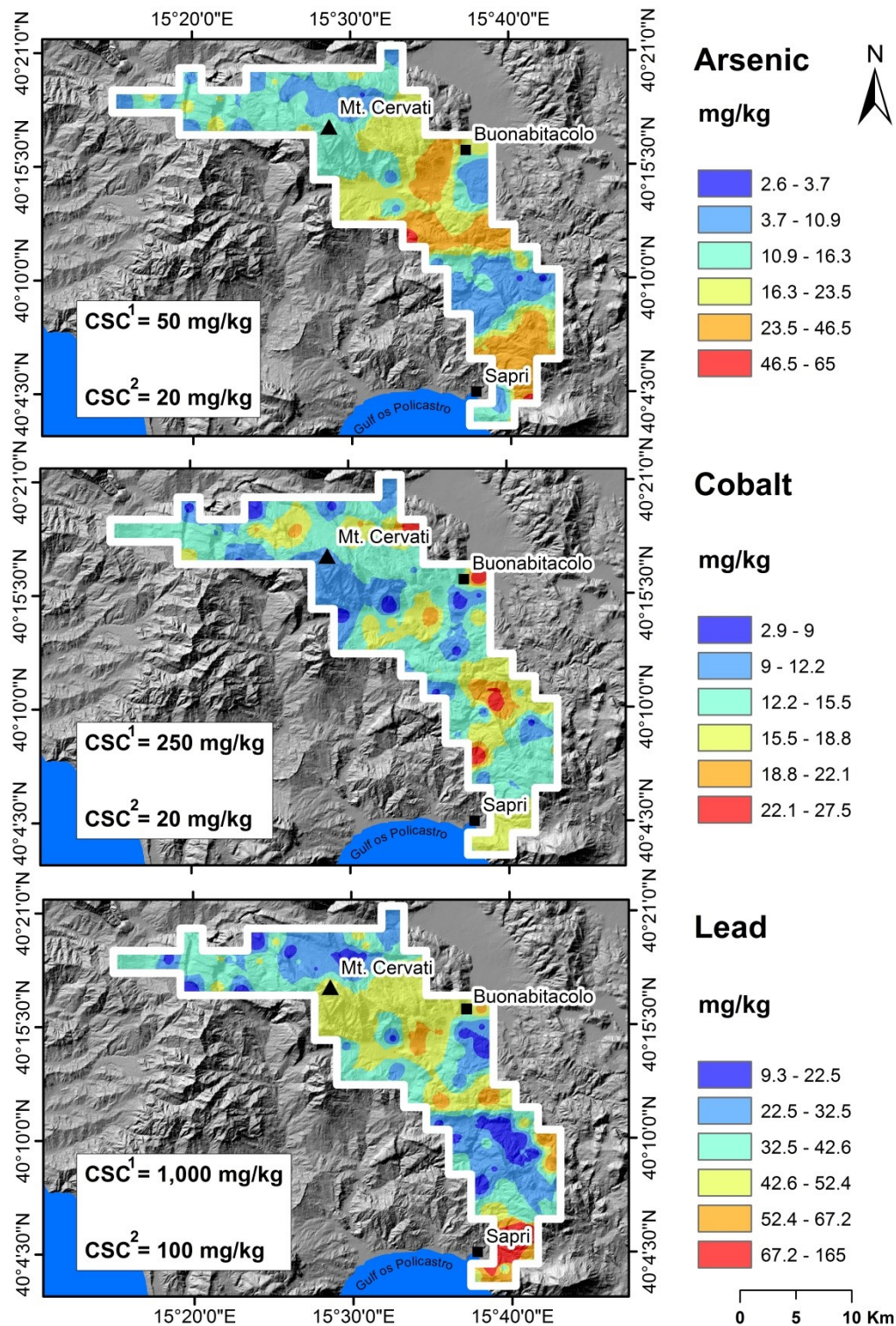


Fig. 4. Geochemical map of As, Co and Pb, compiled using multifractal- IDW interpolation method (GeoDAS). Pixel values have been reclassified by fractal concentration–area (C–A) plot based on the frequency distribution of pixel values. CSC^1 =contamination thresholds established by Italian legislation for industrial/commercial use of soils (Legislative Decree 152/2006). CSC^2 = contamination thresholds established by Italian legislation for residential use of soils (Legislative Decree 152/2006).

For example Be and As concentration are extremely high. Based on previous studies at national scale (Cicchella et al., 2015), it is clear that such high concentrations represent natural background values of these elements exceeding the Italian statutory limit for soil on the whole territory of southern and central Italy. The Pb spatial distribution seems to be controlled mostly by anthropogenic contributes, showing the highest concentration values in correspondence of the main urban centers (e.g., Sapri). Also Co shows concentrations over CSC for residential/recreational use (about 10% of analyzed samples > 20 mg/kg). This type of data analysis does not take into account their compositional nature.

Locally different geochemical contributes could exist. A data analysis CODA-based could be helpful in the interpretation of the phenomena determining the soil geochemistry of the investigated area.

Fig. 5 displays the data distributions of As, Pb and Co into the survey area through an edaplot (combination of histogram, density trace, scattergram and boxplot).

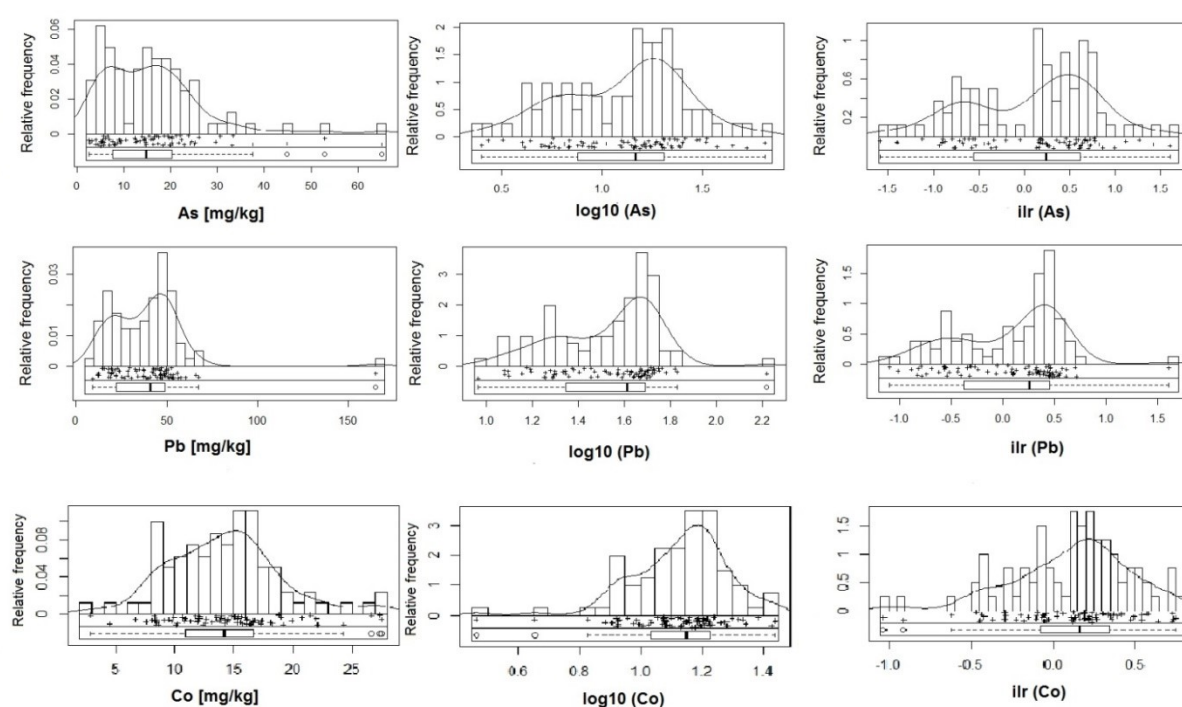


Fig. 5. Edaplots (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) of the raw data, log-transformed data, and ilr transformed data.

This figure shows three different representations of the data: the original concentrations data (left plots), the lognormal- transformed data (middle plots), and the ilr-transformed data (right plots). The graphical data show that the variables do not follow a normal distribution and tend to display a bimodal distribution clearly fitting in the ilr-transformed data plots. As and Pb clearly show two populations in ilr-transformed plots. This observation is clearly displayed on the histogram combined with density traces and the one dimensional scattergram; this might reveal the main activities associated to these elements into the

survey area. The boxplots show several data points in upper inner fence far away from the center of the whisker plotting the raw data; these artefacts tend to disappear or to be highly reduced when log-transformed and ilr-transformed data are used. In addition, these data points are considered as outliers and should be identified and modelled to avoid incorrect results. It proves that compositional data transformation such as ilr transformation solves this problem and moves the composition sample space to the Euclidean Real space R^{-1} , which is failed by the classic statistical data transformation.

The compositional biplot (Fig. 6) based on principal component analysis displays the correlation and relationship between 18 analytical variables in 81 sample points, and the first two principal components extracted. The principal components are presented in a compositional biplot using raw data (Fig. 6, left plot) and ilr coordinates clr back-transformed (Fig. 6, right plot).

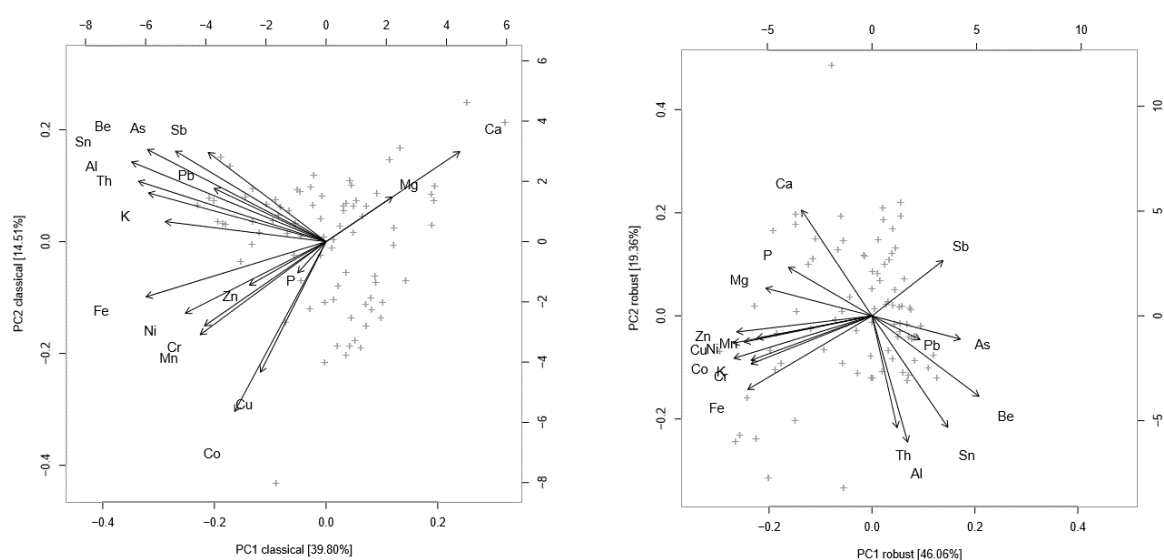


Fig. 6. Biplots for first and second principal components of factor analysis for raw data (classical, left plot) and ilr coordinates clr back-transformed (robust, right plot) of the survey area (Cilento, Southern Italy).

The total variance of initial raw data biplot explains 54.31%, where the first principal component (PC1) accounts 39.80% and the second principal component (PC2) accounts 14.51% (Fig. 6, left plot), whereas the robust biplot based on ilr coordinates clr back transformed produces much more suitable results (Fig. 6, right plot) with PC1 explaining 46.06% and PC2 explaining 19.36% of the compositional variability.

The classical biplot shows clearly three data (variable) elemental associations, whose observations (samples) are widely displayed into the graphic, emphasizing the presence of many outliers or artefacts correlated with Ca and Mg variables. These outliers are widely reduced with the robust compositional biplot. Nevertheless, the classical biplot displays three features of element associations which may be related to the typical lithological background of the survey area:

The Ca and Mg association (A1) presents both variables with high communalities (the length of the ray) with Ca more representative (87.02% of variability) and variables overlapping one to another. This is expected due to the fact that they belong to the same source. These two elements characterize the Mt. Cervati limestones (Trias-Lias), belonging to the Alburno-Cervati-Pollino stratigraphic unity.

Nickel, Fe, Mn, Cr, P, Zn, Cu and Co (A2) form clearly an association, their vertexes laying closer together and also because they tend toward the same direction. Fe and Mn are the most representative elements in A2 association due to the fact that the length their rays (variability) dominate in the group. Zn and P display short vectors which likely characterized poorly presented and seems to be only partially correlated to the others element of A2 association. However, Fe and Mn prove their predominance in the geochemical processes of this group and are likely related to the occurrence of adsorption phenomena. A2 association might be related to the adsorption and coprecipitation effects operated by Fe and Mn oxides and hydroxides occurring mostly in the sedimentary deposits such as marl-sandstone, conglomerate and silico-clastic flysch deposits outcropping in the survey areas.

The third group is represented by Th, Sn, Be, K, Sb, Al, As, Pb and Sb association (A3). Al, Be and Th are elements with the highest variability in this association (A3) due to the fact that the lengths of their rays are larger. Indeed, it's likely related to the fact that these elements are mostly immobile during weathering phenomena of the parental rocks and mostly remain in the residual fraction of soil. Furthermore, this elemental association reflects the influence of pyroclastic deposits from different eruptions of nearby volcanoes (Roccamonfina, Vesuvius, Phlegrean Fields - De Vivo et al., 2010; Buccianti et al., 2015; Mt. Vulture and Aeolian islands - Peccerillo, 2005; Scheib et al., 2014) covering the carbonate (mostly limestone) of the study areas.

The robust PCA biplot “opened” data, displaying better the element associations and reducing the outliers which might influence the results. The length of the rays is linked to the variability of the \ln transformed data back transformed in \ln space, and not to the variables themselves. The robust biplot separates the A3 group of the classical biplot in two real associations which emphasize the chemical structures of the lithological background and the main human activities: Al, Th, Sn and Be association might be correlated to the pyroclastic deposits, whereas, Sb, Pb and As group might be related to the main human activities (vehicle emissions). It appears interesting the display of As ray in the two compositional biplots. In fact the As vertex merges inside the A3 association (controlled by pyroclastic deposits) of the classical biplot.

Instead, for the robust biplot it separates two main groups and is located between these two latter main groups: Th, Be and Sn group related clearly to pyroclastics and Pb and Sb group to anthropogenic sources. This might reflect a mixed contribution: both geogenic and anthropogenic. In addition, the

robust biplot shows P positioned between Ca and Mg revealing that phosphate is sorbed by calcite as commonly occurs in carbonate-rich alkaline soils.

Basically the compositional biplot was able to distinguish particular groups of elements, whereas the factor score maps indicate the incidence of identified associations in each sampled point allowing us to interpret possible correlations to natural peculiarities of the area (e.g. lithologies) and human activities, hence making it possible to discriminate geogenic vs anthropogenic sources of different elements.

Moreover, C-A plot computation considers the compositional nature of geochemical data and their fractal distribution.

Factor analysis was performed on the 18 selected elements (K, Ni, Co, Cr, Fe, Cu, Zn, Sn, Al, Th, Be, Mn, P, Sb, Pb, Ca, Mg and As), using their ilr transformed data clr back transformed A three-factor model, accounting for 70.1% of total data variability, has been chosen and the varimax-rotated factors are reported in Table 4.

Table 4. Varimax-rotated factor (three-factor model) of isometric logratio clr back-transformed variables for 81 topsoil samples from the survey area; bold entries: loading values over |0.50|.

Variables	Factor loadings			Communalities
	1	2	3	
Ca	0.07	-0.80	0.04	0.65
Mg	0.35	-0.51	0.35	0.52
Zn	0.83	-0.18	0.16	0.75
P	0.34	-0.50	0.59	0.68
K	0.70	0.04	0.50	0.69
Th	0.25	0.77	-0.16	0.69
Sn	0.17	0.82	-0.18	0.73
Al	0.24	0.75	0.50	0.82
Be	-0.45	0.64	0.23	0.67
As	-0.38	0.29	-0.64	0.64
Pb	0.22	0.07	-0.55	0.51
Sb	-0.12	-0.25	-0.86	0.82
Ni	0.87	-0.04	0.01	0.75
Cr	0.84	0.19	0.01	0.74
Fe	0.9	0.32	-0.09	0.93
Cu	0.91	-0.04	0.03	0.82
Co	0.95	0.06	0.06	0.91
Mn	0.69	-0.08	0.16	0.55
Eigenvalues	6.71	3.74	2.16	
Total variance in %	37.28	20.8	12.06	
Cum. of total Variance	37.28	58.08	70.14	

The KMO (Kaiser-Maeyer-Olkin) test is over 0.75, indicating a good adequacy of the proposed model. Elements with loadings over the value of 0.50 have been considered to describe the main composition

of each factor. The elements of the associations of the three-factor model, sorted in ascending loading, are:

- F1: Co, Cu Fe, Ni, Cr, Zn, K, Mn
- F2: Sn, Th, Al, Be, - (Ca, Mg, P)
- F3: P, Al, K - (Sb, As, Pb)

The factor score distribution maps were classified using the concentration– area (C-A) fractal method implemented in GeoDAS software.

The F1 elemental association (Co, Cu Fe, Ni, Cr, Zn, K, and Mn) accounts 37.3% of total variance. The factor scores interpolated map (Fig. 7) highlights that the highest values (ranging from 1.71 to 2.39) coincide exactly with the silico-clastic deposits dominated by flysch series.

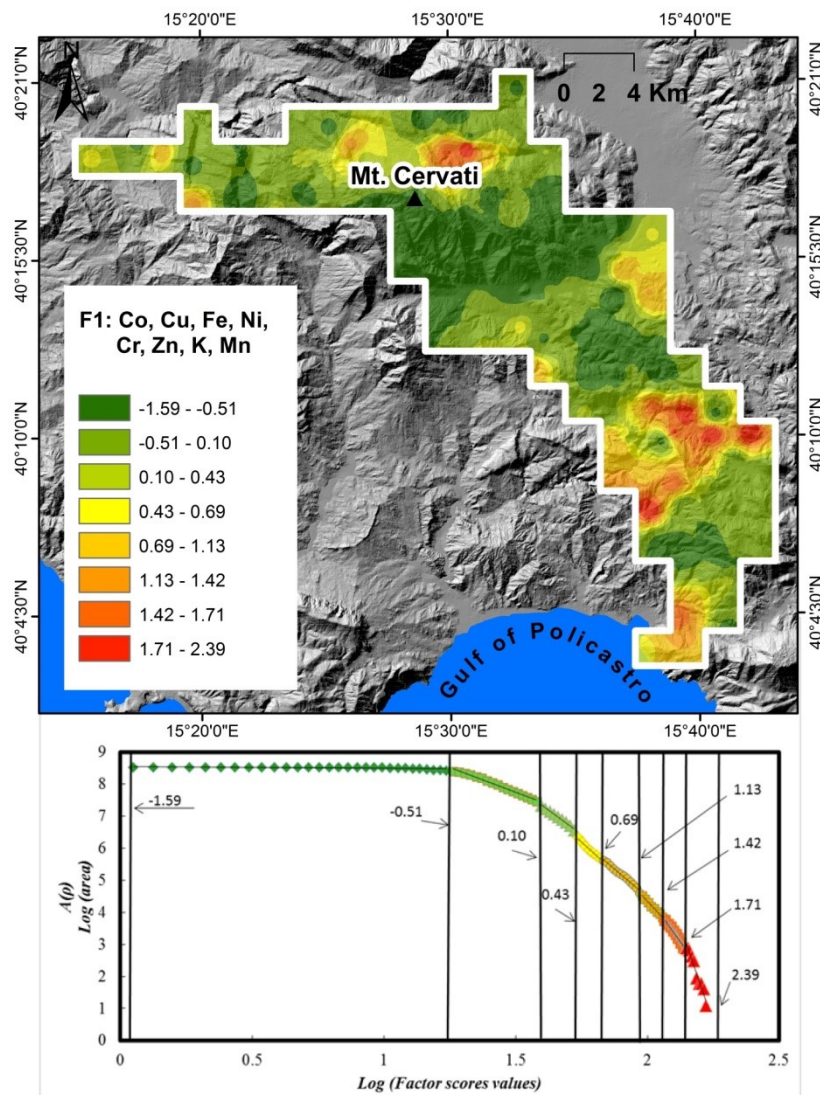


Fig. 7. Map showing the spatial distribution of the Factor 1 of the logratio transformed data; maps are classified through C-A plot (shown below).

Fine size solids with charged substrate such as clay minerals (e.g., smectite/montmorillonite) play an important role in the sorption of trace elements.

Moreover, enrichment of trace elements highlighted by this association reflects adsorption and co-precipitation by Fe and Mn oxy-hydroxides in oxidizing environment. The highest F1 factor score values were found in correspondence of the highest Mn concentrations (up to 1648 mg/kg).

The F2 elemental association [Sn, Th, Al, Be, - (Ca, Mg, P)] accounts a total variability of 20.8% and distinguishes the two groups of elements antithetically correlated (Sn, Th, Al, Be group versus Ca, Mg and P association). Highest factor score values (ranging from 1.74 to 2.21) of this association are located on the slope of the Mt. Cervati (Fig. 8).

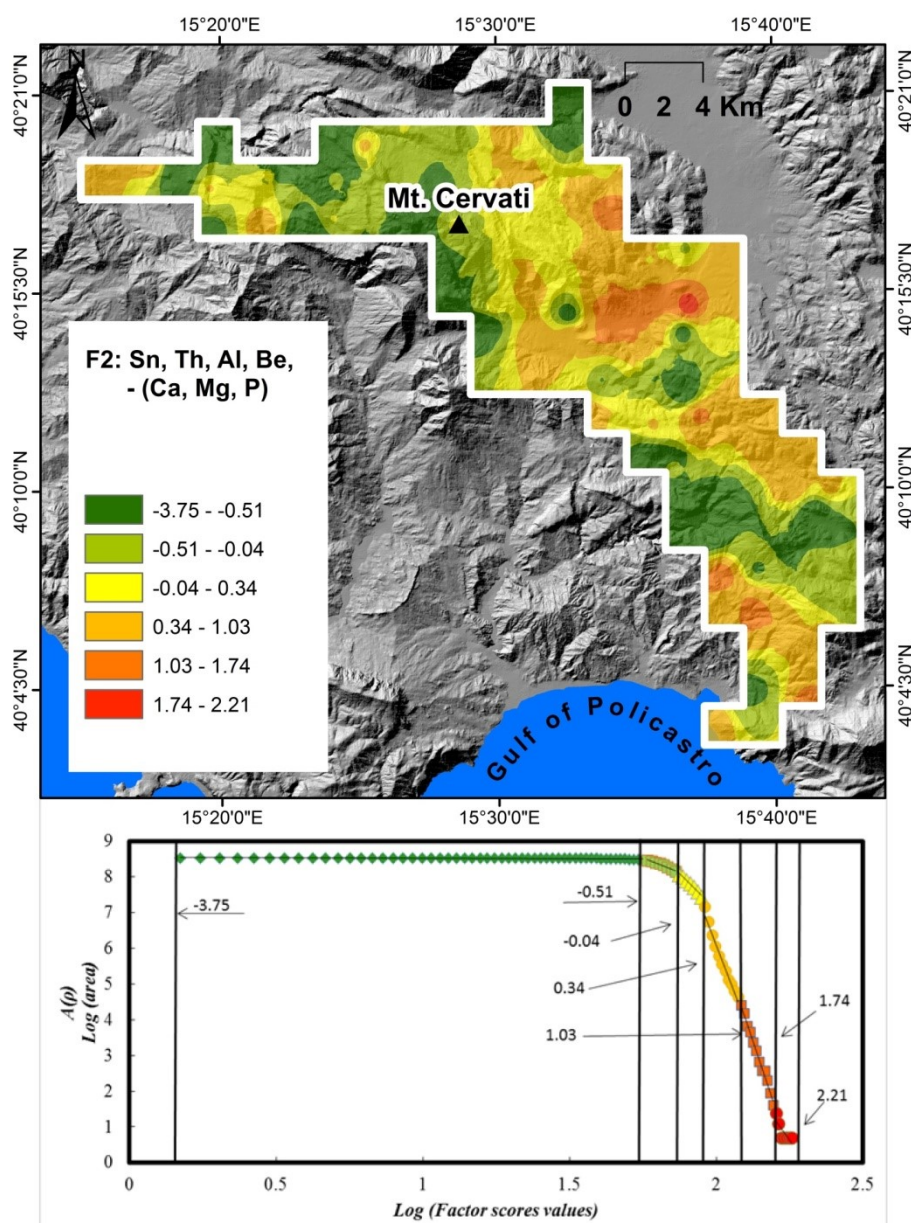


Fig. 8. Map showing the spatial distribution of the Factor 2 of the logratio transformed data; maps are classified through C-A plot (shown below).

F2 maps clearly shows that Sn, Th, Be, Al are strongly controlled by presence of widespread pyroclastics linked to Quaternary volcanic activity of Vesuvius, Phlegrean Fields, Roccamonfina, Mt. Vulture and Aeolian islands (Peccerillo, 2005; De Vivo et al., 2010; Buccianti et al., 2015; Scheib et al., 2014) while Mg-Ca-P association reflects both the dominant carbonate lithology (limestones) of the Mt. Cervati and the carbonate-rich alkaline soils where phosphate is commonly sorbed by calcite. It is interesting to remark that F2 spatial distribution is well explained by the CAI of analyzed soils (Fig. 9). During weathering the proportion of alumina to alkalis would typically increase in the weathered product.

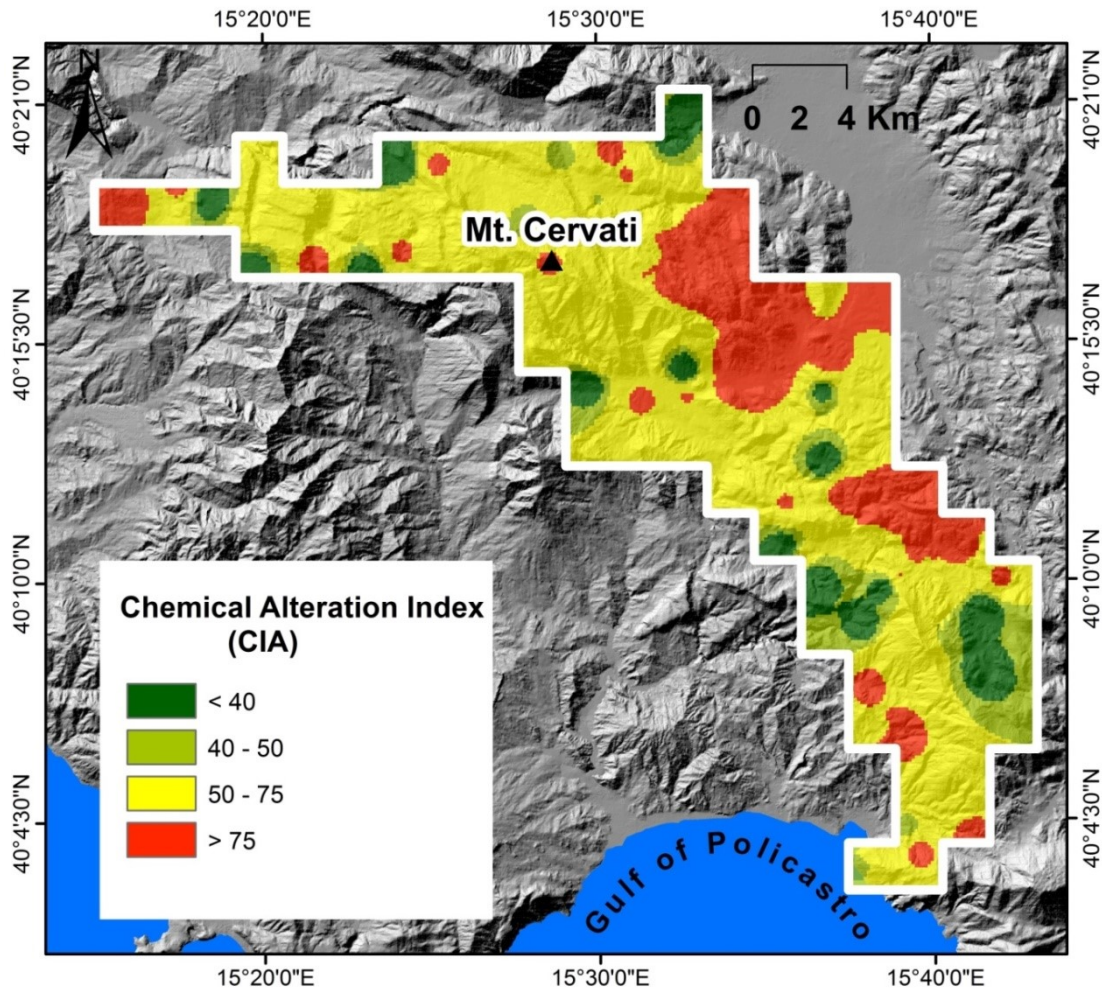


Fig. 9. Spatial distribution of Chemical Alteration Index (CIA). Values < 50 correspond to unweathered soils; > 50 weathered soils (Nesbitt and Young, 1982).

Elements characterized by elevated geochemical mobility, such as Ca and Mg, are negatively correlated with CIA values. Accordingly Sn, Th, Be, Al, characterized by elevated geochemical stability, are found where high-weathered soils occurs. This evidence better clarify the antithetic behaviour of elements describing F2.

The F3 association P, Al, K - (Sb, As, Pb) displays 12.06% of total variance explained (Fig. 10). The highest factor score values (ranging from 0.70 to 2.44) indicative of a higher rate of P, Al, and K, are found mostly in the surrounding mountainous areas, at lower altitude, where soils are impacted by a number of natural and anthropogenic factors: weathering products from both limestone and pyroclastics and more densely populated areas. In any case, high factor scores occur in carbonate and clay-rich alkaline soils that readily incorporate K. limiting its mobility. Potassium may also reflect a contribution from limestone, as impure carbonate can contain up to 6% of K due to the occurrence of clays in non-carbonate fraction (Wedepohl, 1978). Phosphorous, as mentioned earlier is sorbed by calcite.

The antithetic Sb, Pb and As elemental association is strong in correspondence of the most urbanized areas (Sapri, Buonabitacolo and Tortorella) and the main State roads intersections where traffic jams are frequent (Fig. 9). Although Sb does not exceed the contamination thresholds established by Italian legislation for soils (Legislative Decree 152/2006), Pb and As show values above the CSC nearby the urban area of Sapri. It is very interesting to note that the human footprint is clearly recognizable also in environmental contexts characterized by a low anthropogenic pressure (the study area mostly falls in a World Heritage Park).

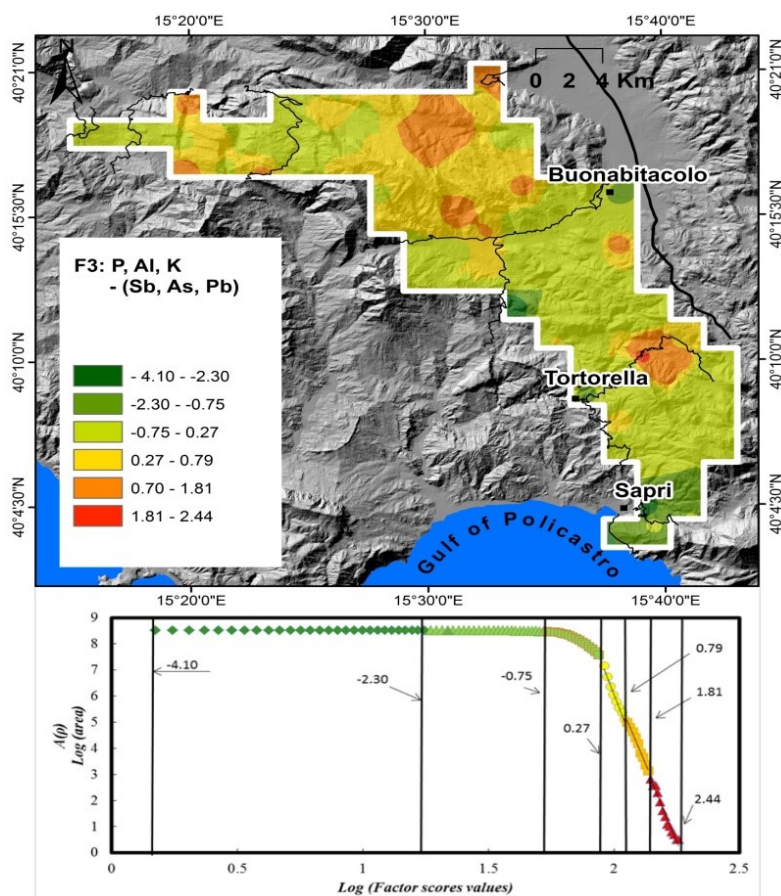


Fig. 10. Map showing the spatial distribution of the Factor 3 of the logratio transformed data; maps are classified through C-A plot (shown below).

6. Remarks and conclusion

The Cilento - Vallo di Diano area (Campania region, Southern Italy) is characterized by several PTEs exceeding the contamination thresholds (CSC) established by Italian legislation for soils (Legislative Decree 152/2006).

This study represents a geochemical follow-up of the mentioned area, in order to assess the distribution pattern and the behaviour of the geochemical anomalies, performing a methodology, which takes into account both compositionality and fractal distribution of geochemical data.

The ilr normalization moves the compositional data from the sample space to the Euclidean real space preserving the orthonormal basis. The frequency based methods (univariate) such as Edaplots of raw, log-transformed and ilr-transformed data were used to depict data distribution; the ilr-transformed variables show a clear bimodal distribution for As and Pb. This type of distribution tends to disappear when raw data and log-normal data are plotted. In fact data tend to be normalized when ilr-transformed data are used (Filzmoser et al., 2009b).

Moreover, the main benefit of ilr-transformation, compared to raw data and log-normal data, is the enhancement of anomalous multi-element associations reflecting the geogenic, anthropogenic or mixed contribution affecting the geochemical dataset.

The classical biplot based on raw data shows the main geological feature of the study area identifying three lithological complexes such as:

- (1) a first association (Ca and Mg) reflecting limestones and dolostone occurrence of the Mt. Cervati;
- (2) a second association (Ni, Fe, Mn, Cr, Cu, Zn and Co) reflecting the coprecipitation effect of Mn and Fe oxides-hydroxides in flysch and arenaceous material of the survey area;
- (3) a third association (Th, Sn, and Be) related to the pyroclastic coverings.

The robust PCA based on a robust covariance estimator like the MCD, clearly shows the effect of the data opening. The variables in robust biplot (ilr coordinates clr back-transformed) are more evenly distributed across the component space, making easy the recognition of element associations characteristic of different geochemical (natural, anthropogenic or mixed) processes. This robustification clearly displays additional phenomena, which affect the natural features of the survey area. In fact, Sb, Pb and As are well discriminated in the robust biplot and may be related to the influence of anthropogenic activities (vehicle emission).

The combination of C-A fractal model and factor analysis on log-transformed variables, allow us to recognize the geological features and the geochemical processes controlling the presence and distribution of chemical elements. In our survey area soil alteration phenomena could mask the nature of parental rock. The factor scores maps highlight an antithetic behaviour of many elements. Elements

characterized by elevated geochemical mobility, such as Ca and Mg, are negatively correlated with CIA values. Instead, Sn, Th, Be, Al are characterized by elevated geochemical stability; for this reason they are found where high-weathered soils occurs. In the latter, surprisingly, also high F3 factor scores (with elements such as P and K) occur likely due to carbonate and clay-rich alkaline soils. The presence of elements such as Co, Cu Fe, Ni, Cr, Zn, K, Mn is mainly controlled by terrigenous flysch deposition. Arsenic, Pb (exceeding the CSC) and Sb association mainly occur in correspondence of urban areas and where traffic jams are frequent. It is very interesting to note how the human footprint (As, Sb Pb) is clearly recognizable also in environmental contexts characterized by low anthropogenic pressures (the study area mostly falls in a World Heritage Park).

According to our results and observations, this method demonstrate to be a useful tool to distinguish different processes controlling the elemental geochemical distribution in our study area, highlighting additional phenomena which would normally be masked.

Chapter 3

Environmental geochemistry of organic pollutants: Human and Ecological health risk assessment

Section 3.1

Source patterns and contamination level of polycyclic aromatic hydrocarbons (PAHs) in urban and rural areas of Southern Italian soils

Section 3.2

Status, sources and contamination levels of organochlorine pesticide residues in urban and agricultural areas: a preliminary review in central-southern Italian soils

Section 3.1

Source patterns and contamination level of polycyclic aromatic hydrocarbons (PAHs) in urban and rural areas of Southern Italian soils

This section has been published in:
Journal of Environmental Geochemistry and Health
Available online, 29 June 2018

Source patterns and contamination level of polycyclic aromatic hydrocarbons (PAHs) in urban and rural areas of Southern Italian soils

Abstract

Polycyclic aromatic hydrocarbons (PAHs) are a group of persistent organic pollutants. They have been identified as a type of carcinogenic substance and are relatively widespread in environment media such as air, water and soils, constituting a significant hazard for human health. In many parts of the world, PAHs are still found in high concentrations despite improved legislation and monitoring, and it is therefore vital defining their profiles, and assessing their potential sources. This study focused on a large region of the south of Italy, where concentration levels, profiles, possible sources and toxicity equivalent quantity (TEQ) level of sixteen PAHs were investigated. The survey included soils from five large regions of the south of Italy: 80 soil samples (0-20 cm top layer) from urban and rural locations were collected and analysed by gas chromatography-mass spectrometry. Total PAHs and individual molecular compounds from the US Environmental Protection Agency priority pollutants list were identified and measured. Results showed that 16 PAHs varied significantly in urban and rural areas, and different regions presented discordant characteristics. Urban areas presented concentrations ranging from 7.62 to 755 ng g⁻¹ (mean = 84.85 ng g⁻¹), whilst rural areas presented ranges from 1.87 to 11,353 ng g⁻¹ (mean = 333 ng g⁻¹). Large urban areas, such as Rome, Naples and Palermo, exhibited high PAHs total concentration, but high values were also found in rural areas of Campania region. Different PAHs molecular ratios were used as diagnostic fingerprinting for source identification: LWMPAHs/HWMPAHs, Fluo/(Fluo + Pyr), BaA/(BaA + Chr), Ant/(Ant + Phe) and IcdP/(IcdP + BghiP). These ratios indicated that PAHs sources in the study area were mainly of pyrogenic origin, i.e. mostly related to biomass combustion and vehicular emission. On the other hand, values in Sicilian soils seemed to indicate a petrogenic origin, possibly linked to emissions from crude oil combustion and refineries present in the region. Finally, results allowed to calculate the toxicity equivalent quantity (TEQBAP) levels for the various locations sampled, highlighting that the highest values were found in the Campania region, with 661 and 54.20 ng g⁻¹, in rural and urban areas, respectively. These findings, which could be linked to the presence of a large solid waste incinerator plant, but also to well-documented illegal waste disposal and burning, suggest that exposure to PAH may be posing an increased risk to human health in some of the studied areas.

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) are diffuse persistent organic pollutants (POPs) that can be found in different environmental media, including air, water and soil. They are human carcinogens, mutagens and are toxic to all living organisms, making them a group of compounds of public concern, which are becoming increasingly studied and monitored in many areas of the world (IARC 1983; Hwang et al. 2003; Nadal et al. 2004; Vane et al. 2014). PAHs are primarily formed through the incomplete combustion of carbon containing fuels such as wood, coal, diesel, fat and tobacco, and most sources of PAHs are anthropogenic, arising from industrial emissions, solid waste incineration and vehicular emissions among others (Dong and Lee 2009). Sixteen US Environmental Protection Agency (EPA) priority PAHs are classified in two main groups of compounds related to the number of aromatic rings: low molecular weights PAHs (LMWPAHs) with 2–3 aromatic rings (naphthalene, acenaphthylene, acenaphthene, fluorene, phenanthrene and anthracene) and high molecular weight PAHs (HMWPAHs) with 4–6 aromatic rings such as fluoranthene, pyrene, benzo[a]anthracene, chrysene, benzo[b]fluoranthene, benzo[k]fluoranthene, benzo[a]-pyrene, indeno[1,2,3-cd]pyrene, dibenzo[a,h]anthracene, and benzo[g,h,i]perylene. PAHs from a petrogenic source are formed predominantly with those of low molecular weights, whilst the PAHs from a pyrogenic source generally have high molecular weights (Soclo et al. 2000). Once formed by the mechanisms of partial combustion, PAHs can be found in different media. In particular, soil is considered an important media to quantify PAHs patterns due to its physicochemical properties that allow PAH compounds to be held in soil matrices (Means et al. 1980). PAHs are slightly or completely insoluble in water, and they are adsorbed on soil particles, particularly on soil organic matter. Hence, the physical–chemical properties of soils are responsible for the retention of PAHs in soil matrices. The organic carbon content, the hydrophobicity of soil organic matter and soil texture were estimated to be the most significant parameters controlling the environmental availability of PAHs (Albanese et al. 2015a). Furthermore, some studies (Menzie et al. 1992; Nadal et al. 2004) have demonstrated that the amount of human exposure to PAHs through soils was higher than through air or water. As they exist in different forms with a different degree of toxicity, it is important to characterise individual PAHs compounds as much as possible. However, given their number and variety, often their ratios can be a more effective diagnostic tool to identify potential source patterns and quantify the amount of PAH pollution for a specific area (e.g. Pandey et al. 1999; Yunker et al. 2002; Hwang et al. 2003). One of the most widespread computations used is that involving the LMWPAHs/HMWPAHs ratio introduced by Budzinski et al. (1997) who fixed a value below 1 for pyrogenic source and above 1 for a petrogenic PAHs fingerprint. Since the introduction of this method, other authors have developed alternative takes by using individual PAH compounds, in particular low

and high molecular weight groups, to highlight their main source patterns (Hwang et al. 2003, Yunker et al. 2002; Tobiszewski and Namiesnik 2012). Recent development of analytical techniques has seen an increase in the number of studies focusing on individual PAHs compounds, and several studies (Zhang et al. 2006; Albanese et al. 2015a; Islam et al. 2017) have focused on topsoil PAHs occurrence and concentration, helping to shape and informing government policy for human and ecological safety. For example, Italian environmental law (D. Lgs. 152/2006) establishes threshold values that regulate the mitigation of the PAHs in soil media.

This legislation fixes different PAHs concentration values based on the type of PAHs and the land use (e.g. residential and industrial areas). Regulations can guide efforts and resources for reclamation and more detailed monitoring, contextualising interpretations in line with risk-based approaches. However, much needs to be done to establish baselines and understand the mechanisms of these contaminants' movement and availability in the environment. In particular, in southern Italy there exist several potential anthropogenic sources of PAHs such as petroleum exploitation districts, biomass combustion plants, vehicular emissions and residential wood combustion which can all constitute a source of PAHs compounds and contribute to their concentrations in soils and other media. These have been only recently studied, and mainly at a local, small scale, whilst a regional baseline approach has not been carried out yet.

The present study will focus on the 16 (EPA) PAHs priority compounds to carry out a regional survey in southern Italy. PAHs will be characterised in soils of several urban and rural locations to assess their spatial distribution, their potential sources and pathways their level of toxicity.

This study is important because it will constitute the first regional survey carried out in Italy and can be considered a first stepping stone towards a more detailed and meaningful investigation on potential sources and levels of PAHs in the region. It is anticipated that this study will contribute to build a baseline for PAHs characterisations in urban and rural areas of southern Italy. Follow-up studies should be expected in areas where high PAHs concentration levels (i.e. contamination) were found, with a larger number and higher density of (soil and air) samples in each affected location.

2. Materials and methods

2.1. Study area

The survey area included five administrative regions (Lazio, Campania, Basilicata, Calabria and Sicily) of the south of Italy. The total area extends to approximately 81,054 km² with 19.38 million inhabitants, mostly grouped in urban areas. The overall area is characterised by a principal mountain range, the Apennine chain, which presents very specific geological and morphological features (Fig. 1a).

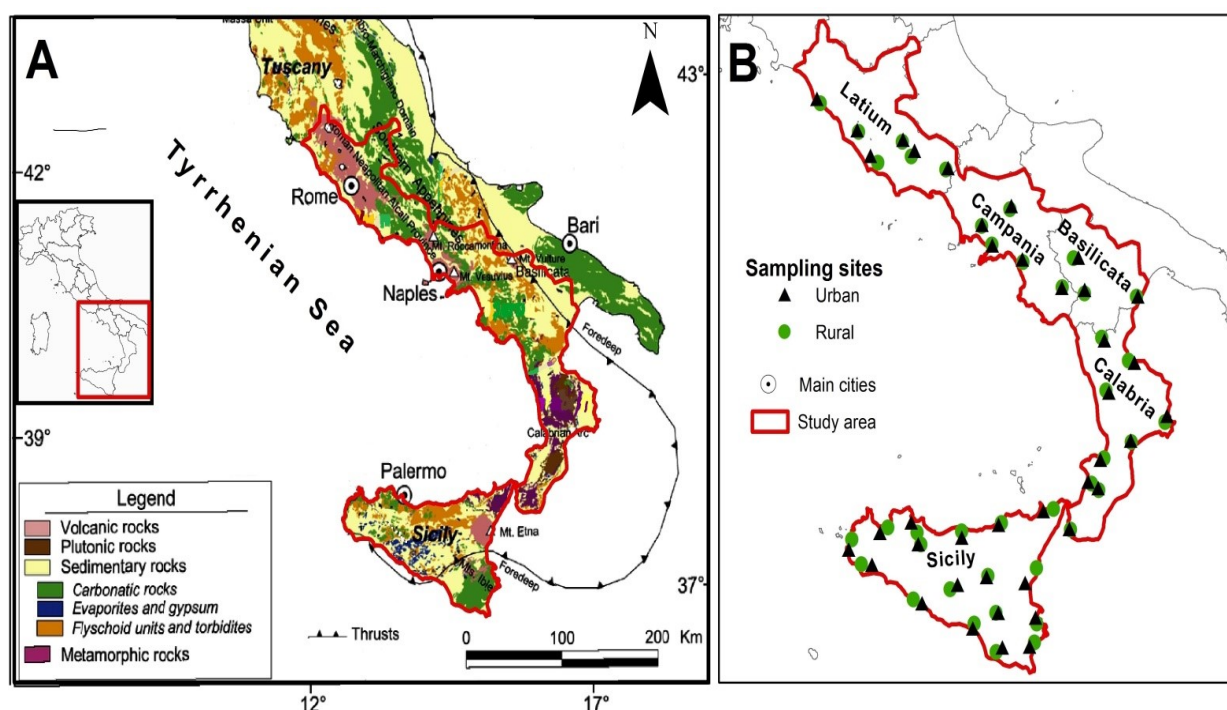


Fig. 1. (A) Simplified Geology of the study area, modified after Doglioni and Flores (1997). (B) Survey area with samples sites.

The main geological features of the study area are a result of the tectonic and orogenic activities related to the nearby boundary between the Eurasian and the African Plates. The area is characterised by active volcanism of different origin, such as that related to the Vesuvius, the Phlegraean Fields, Ischia island, Pontine Islands and Roccamonfina in the Campania region, and the Etna and Aeolian islands in the Northeast of Sicily. Such volcanism can be considered a potential (though limited) source of PAHs during volcanic eruptions (Kozak et al. 2017). From the land use point of view, the study region is devoted to intensive agriculture activities such vineyards, olive plantations—mostly in hilly areas—citrus fruits, seasonal crops and greenhouse products (tomatoes, potatoes, aubergines, peppers and peas) which represent major resources for the local economy (Albanese et al. 2007). In addition to this, large industrial works are also present in the region, processing raw materials of various origins (e.g. petroleum plants, biomass and alloy). Some of these industrial activities can be linked to ‘petrogenic’ sources of organic pollutants. Some good examples can be found in: (1) the Priolo and Gela districts (Sicily), where the largest petroleum field and refinery in Europe can be found (DPCM 1990); (2) Alta Val d’Agri (Basilicata), which is the location of the largest Italian inland oil field; (3) Crotone (Calabria) and Milazzo (Sicily), the sites of two very large petroleum refineries. An industrial activity that can be linked to potential ‘pyrogenic’ sources of organic pollutants is that related to energy (and energy-from waste) generation, which is widely diffused in the study area. Some good examples can be found in:

- (1) Latium, with the thermal and coal power plants in Civitavecchia and the solid municipal waste incinerator plant in Frosinone;
- (2) in Campania, with various large industrialised areas within the metropolitan area of Naples, and with the urban solid waste Incinerator plant of Acerra; (3) in Basilicata (Potenza and Melfi) and Calabria (Gioia Tauro), where other incinerators plants can be found (Fig. 2).

2.2. Description of sampling locations

The sampling campaign took place from early April to end of September 2016, with the aim of taking the most representative soil samples in urban and rural areas throughout five administrative regions (Latium, Campania, Basilicata, Calabria and Sicily) of southern Italy (Fig. 2). In each region, the main urban areas and the nearest rural areas, where most of the land is devoted to agricultural activities, were selected. The sampling site characterisation and selection was performed using geographical information systems (GIS) data of the industrial and agricultural activities and land use of the study area (ISPRA 2014; ISTAT 2016) as well as using satellite images (Google Earth professional, 2016 version).



Fig. 2. Land use and main industrial activities in the study area. This industrial repertory is based on the ISPRA (2014) annual report.

2.3. Sample collection and materials

A total of 80 soil samples were collected for polycyclic aromatic hydrocarbons (PAHs) with a nominal density of 2 samples (in urban and rural areas) in each 2500 km² (Fig. 1b). The sampling procedure followed the Geochemical Mapping of Agricultural and Grazing Land Soils (GEMAS) sampling procedure described by Reimann et al. (2005). All the samples were collected using a stainless steel

scoop and were kept in labelled glass bottles and directly stored in ice boxes to minimise the losses caused by volatilisation and initial degradation of the organic compounds (Albanese et al. 2015a). Each topsoil sample (from 0 to 20 cm) was made by homogenising five subsamples at the corners and the centre of a 100 m squared, collecting approximately 1.5 kg in total. The homogenised soil samples were sieved using a < 2 mm mesh sieve after removing stones, detritus and residual roots. Finally, composite samples were stored at -4 °C in the environmental geochemistry laboratory of the University of Naples Federico II (Italy) until instrumental analysis. For each sampling site pH, moisture content and electric conductivity of the soil were measured. Records of the land use, main industrial works and any other human activity observed in the proximity of the sampling locations were recorded at each site, where geographical coordinates were uploaded by global positioning systems (GPS). These attributes would form the dataset subsequently used for the spatial analysis and representation by GIS.

2.4. Sample preparation and analytical procedures

Samples were analysed for PAHs content at the Key Laboratory of Biogeology and Environmental Geology of Ministry of Education, China University of Geosciences, Wuhan, China. For this study, the target analytes were the 16 US EPA priority PAH compounds: naphthalene(Nap), acenaphthylene (Acy), acenaphthene (Ace), fluorine (Flu), phenanthrene (Phe), anthracene (Ant), fluoranthene (Fluo), pyrene (Pyr), benzo[a]anthracene (BaA), chrysene (Chr), benzo[b]fluoranthene (BbF), benzo[k]fluoranthene (BkF), benzo[a]-pyrene (BaP), indeno[1,2,3-cd]pyrene (IcdP), dibenzo[a,h]anthracene(DahA), and benzo[g,h,i]perylene (BghiP). Ten grams of soil sample was weighed and injected with PAH surrogates (naphthalene-d8, acenaphthene-d10, phenanthrene-d10, chrysene-d12 and perylene-d12) and Soxhlet extracted with dichloromethane (DCM) for 24h. The extracts were treated with activated copper granules to remove elemental sulphur, concentrated and solvent exchanged to n-hexane and further reduced to 2–3 ml by a rotary evaporator (Heidolph 4000, Germany). A 1:2 (v/v) alumina/silica gel column (450 °C muffle drying for 4 h, both 3% deactivated with H₂O before using) was used to clean up the extracts, and PAHs were eluted with 70 ml of DCM/hexane (2:3). The eluate was then reduced to 0.2 ml under a gentle stream of nitrogen. A known quantity of hexamethylbenzene was added as an internal standard for PAHs analysis prior to instrumental quantitation for the PAHs.

PAHs were analysed using GC-MS (Agilent 6890 N/5975 MSD) coupled with a HP-5972 mass selective detector operated in the electron impact mode (70 eV) installed with a DB-5 capillary column (30 m 9 0.25 mm diameter, 0.25 µm film thickness). Helium (99.999%) was used as the GC carrier gas at a constant flow of 1.5 ml min⁻¹. An 1 µl concentrated sample was injected with splitless mode. The chromatographic conditions were as follows: injector temperature 270 °C; detector temperature 280 °C;

oven temperature initially at 60 °C for 5 min, increased to 290 °C at 3 °C min⁻¹, and held for 40 min. Chromatographic peaks of samples were identified by mass spectra and retention time.

2.5. Quality control

Procedure types used for quality assurance and quality control (QA/QC) were as follows: method blank control (procedural blank samples), parallel sample control (duplicate samples), solvent blank control, and basic matter control (US EPA, 2002). In order to ensure the validity of the analyses during the experiment, different reagents and procedures were used:

1. One thousand Nanograms (ng) of naphthalene-D8, acenaphthene-D10, phenanthrene-D10, chrysene-D12, and perylene-D12 were used as recovery surrogates, and 1000 ng hexamethylbenzene was added in extracts as the internal standard substance. The spiked recoveries of PAHs using composite standards were 79.9±14.5 % for naphthalene-D8, 74.2±9.4 % for acenaphthene-D10, 91.5±11.6 % for phenanthrene-D10, 87.1±8.5 % for chrysene-D12, and 89.2±11.0 % for perylene-D12, respectively.
2. An internal standard method was used for quantification: a six-point calibration curve was established according to the results from the PAH-16 standard reagents with concentration of 10, 5, 2, 1, 0.5, and 0.2 mg.l⁻¹. For PAHs, the target compounds were identified on the basis of the retention times and selected quantitative ion.
3. During the pre-treatment, a procedural blank and a parallel sample consisting of all reagents was run to check for interference and cross contamination in every set of samples (about 16 samples). Only low concentrations of few target compounds can be detected in procedural blank samples. For more than 96 % of target compounds in parallel samples, the relative error (RE, %) of concentrations are less than 50 %, which is acceptable for Specification of Multi-purpose Regional Geochemical Survey and Guidelines for sample analysis of Multi-purpose Regional Geochemical Survey recommended by China Geological Survey (DD2005-1 and DD2005-3);
4. During the GC-MS analysis period, a solvent blank sample and a PAH-16 standard reagent with concentration of 5 mg l⁻¹ and 100µg l⁻¹ were injected every day before analysing the soil samples. The target compounds were not detectable in the solvent blank samples.
5. Multi-injections were used for precision or accuracy. The samples of different concentrations were injected continually for ten times, and the relative standard deviation was calculated. RSD for all the target compounds ranged from 3.2 to 7.9 %. The final concentrations of PAHs in all samples were corrected according to the recovery of the surrogates and the results of blank samples were subtracted.

2.6. Statistical computations

Univariate and multivariate analyses were carried out on the 16 PAHs through descriptive statistics and compositional principal factor analysis modelling, which helped displaying the variation of these compounds and their main correlations in the survey area. Computations and graphical representations were implemented by mean of the open source statistical software R (Templ et al. 2011).

2.7. Source apportionment

2.7.1. PAHs diagnostic ratios

Whilst several PAHs diagnostic ratios are available in the literature, (e.g. Katsoyiannis et al. 2007; Ravindra et al. 2008), in this study, four specific PAH molecular ratios were used for the identification of PAHs pollution sources: LMWPAHs/HMWPAHs, Fluo/(Fluo + Pyr), IcdP/(IcdP + BghiP) and BaA/(BaA + Chr) (Figs. 5, 6; Table 1).

Table 1. Compilation of PAH ratios and molecular markers for source diagnosis.

Compounds, compounds ratios	Values, ranges	Sources	References
<i>LMWPAHs/HMWPAHs</i>	≤ 1	Pyrogenic combustion	Budzinski et al., (1997)
	≥ 1	Petrogenic source	Budzinski et al. (1997)

<i>Fluo/(Fluo + Pyr)</i>	< 0.40	Petroleum/petrogenic source	Yunker et al. (2002)
	0.4 - 0.5	Fossil fuel combustion	Yunker et al. (2002)
	> 0.5	Biomass and coal combustion	Yunker et al. (2002)

<i>BaA/(BaA+Chr)</i>	< 0.20	Petroleum/petrogenic source	Yunker et al. (2002)
	0.2-0.35	Coal combustion	Yunker et al. (2002)
	> 0.35	Traffic emission	Yunker et al. (2002)

<i>IcdP/(IcdP+ BghiP)</i>	< 0.20	Petrogenic source	Tobiszewski and Namieśnik (2012)
	0.20 - 0.50	Fuel combustion	Tobiszewski and Namieśnik (2012)
	> 0.50	Coal and biomass combustion	Tobiszewski and Namieśnik (2012)

In particular, the reasons why these were chosen are that: the ratio LMWPAHs/HMWPAHs ≤ 1 corresponds to pyrogenic sources, and > 1 for petrogenic sources (Budzinski et al. 1997); the ratio Fluo/(Fluo + Pyr) < 0.4 has been shown to indicate petroleum source, between 0.4 and 0.5 implies fossil fuel combustion, and a ratio > 0.5 is the characteristic of biomass and coal combustion (Yunker et al. 2002). IcdP/(IcdP + BghiP) < 0.2 is an indication of petroleum sources, whilst that between 0.2 and 0.5

indicates that the PAHs usually derive from petroleum combustion (liquid fossil fuel, vehicle and crude oil combustion) and $\text{IcdP}/(\text{IcdP} + \text{BghiP}) > 0.5$ strongly indicates the contribution of coal, grass and wood combustion (Tobiszewski and Namieśnik 2012). Yunker et al. (2002) implemented the $\text{BaA}/(\text{BaA} + \text{Chr})$ ratio and the value < 0.2 to mark a petroleum/petrogenic source, whilst that between 0.20 and 0.35 is linked to a combustion and > 0.35 is related to a traffic emission.

2.7.2. Compositional multivariate computation: factor analysis

R-mode factor analysis, a type of multivariate statistics, was chosen to explain the correlation structure of the 16 EPA PAHs compounds (variables) using a smaller number of factors (Reimann et al. 2002). This methodology has been shown to successfully correlate the PAHs distribution to their main hypothetical origins (Albanese et al. 2015a; Islam et al. 2017). To minimise and/or eliminate the presence of outliers and spurious correlation (Pawlowsky-Glahn and Buccianti 2011), isometric log-transformed data (ilr) are recommended in this type of multivariate analysis (Filzmoser et al. 2009). In order to facilitate the interpretation of results, varimax rotation was used, since it is an orthogonal rotation that minimises the number of variables that have high loadings on each factor, simplifying the transformed data matrix and assisting interpretation (Reimann et al. 2002). The different factors obtained were studied and interpreted in accordance with their presumed origin (petrogenic, pyrogenic or mixed) (Table 4, Fig. 7). The factor score values were mapped at each sample site using GIS software GeoDAS (Cheng et al. 2001) and ArcGIS (ESRI 2012). GeoDASTM was used to produce interpolated geochemical maps of the factor scores by means of the multifractal inverse distance weighted (MIDW) algorithm (Cheng et al. 1994; Lima et al. 2003). MIDW is one of the most widely used interpolation methods on geochemical data because it preserves high-frequency information, retains local variability taking into consideration both spatial association and local singularity (Cheng et al. 1994, 2001; Lima et al. 2003). Singularity is an index representing the scaling dependency from multifractal point of view, which characterises how the statistical behaviour of a spatial variable changes as the measuring scale changes (Cheng et al. 1994). Spatial association represents a type of statistical dependency of values at separate locations, and its indexes (e.g. covariance, autocorrelation and semivariogram) have been used to characterise the local structure and variability of surfaces (Cheng et al. 1994). During interpolation and mapping of geochemical variables, both spatial association and scaling are taken into account. Despite the low density, interpolation is still a valid tool at regional level, as shown, for example, by the production of European geochemical atlases, which have used similar techniques (Reimann et al. 2012; Ottesen et al. 2013; Albanese et al. 2015b). The concentration–area (C–A) fractal method (Cheng et al. 1994) was applied to set the factor score intervals (Thiombane et al.

2018) in interpolated surface images generated by the MIDW method, and ArcGIS™ software was used for the graphical presentation of the results (Figs. 8, 9, 10).

2.8. Characterisation and toxicity assessment of PAHs

For each PAH compound, the toxicity equivalency factor (TEF) was established to allow measuring its relative carcinogenicity (EPA 1984; Nisbet and Lagoy 1992). Among the 16 EPA PAHs, seven of them, including BaA, Chr, BaP, BbF, BkF, IcdP and DahA, present high toxic and carcinogenic effects. In particular, benzo[a]-pyrene (BaP) is considered as one of the most toxic PAHs and it has been used to quantify the relative toxicity of others PAHs (Nisbet and Lagoy 1992), also because it is the only compound with sufficient toxicological data to derive carcinogenic factors among all other potentially carcinogenic PAHs. The toxicity of soil can be measured using the BaP toxic equivalent quantity (TEQ_{BaP}) (Nadal et al. 2004) for each sampling site using the equation described below:

$$TEQ_{BaP} = \sum_{i=1}^7 TEF \times C_{PAHi} \quad (1)$$

where TEQ_{BaP} is the toxic equivalent quantity of ithPAH, the TEF of BaA, Chr, BaP, BbF, BkF, IcdP and DahA corresponds to 0.1, 0.01, 1, 0.1, 0.1, 0.1 and 1, respectively, and C_{PAHi} is the concentration of the ith PAH in the soil.

3. Results and discussions

3.1. Variety PAHs concentrations in the survey area

Table 2 shows the descriptive statistic of the 16 EPA PAHs compounds in soils found on the sample locations of the studied area.

The total concentration of the 16 PAHs in urban and rural area ranged from 7.62 to 755 ng g⁻¹ with a mean value of 84.85 ng g⁻¹, and from 1.87 to 11,353 ng g⁻¹ with a mean of 333 ng g⁻¹, respectively. The spatial distribution of the P16 PAHs concentrations in the survey areas can be seen in Fig. 3, which compares urban and rural areas using proportional thematic mapping.

Table 2. Descriptive statistic of the 16 US EPA PAHs compounds from the survey area.

PAHs (ng/g)	Abbreviation	DL *	TEF**	Urban				Rural			
				Min	Mean	Median	Max	Min	Mean	Median	Max
Naphthalene	Nap	0.013	0.001	0.39	2.49	2.01	9.83	0.33	3.21	2.05	25.13
Acenaphthylene	Acy	0.028	0.001	0.08	0.60	0.20	8.07	0.03	1.41	0.21	22.78
Acenaphthene	Ace	0.047	0.001	0.04	0.24	0.11	2.40	0.03	0.79	0.11	25.20
Fluorene	Flo	0.009	0.001	0.12	0.61	0.57	1.49	0.11	1.21	0.53	25.05
Phenanthrene	Phe	0.001	0.001	0.85	6.74	3.34	47.32	0.40	17.87	2.80	570
Anthracene	Ant	0.005	0.01	0.04	0.77	0.19	8.57	0.01	3.78	0.14	131
Fluoranthene	Fluo	0.003	0.001	0.44	10.43	1.68	108	0.13	25.22	1.57	829
Pyrene	Pyr	0.003	0.001	0.40	10.74	1.60	113	0.11	30.98	1.61	1,059
Benz[a]anthracene	BaA	0.002	0.1	0.25	7.27	1.04	93.30	0.19	26.86	0.98	950
Chrysene	Chr	0.004	0.01	0.41	8.58	1.84	93.15	0.10	33.04	1.85	1,101
Benzo[b]fluoranthene	BbF	0.004	0.1	0.22	15.05	2.75	151	0.07	86.58	2.43	3,055
Benzo[k]fluoranthene	BkF	0.004	0.1	0.19	5.09	0.99	51.38	0.05	19.38	0.84	633
Benzo[a]pyrene	BaP	0.001	1	0.12	10.47	1.32	112	0.08	70.14	1.14	2,510
Indeno[1,2,3-c,d]pyrene	IcdP	0.003	0.1	0.04	2.64	0.42	33.51	0.01	7.91	0.31	226
Dibenzo[a,h]anthracene	DahA	0.003	1	0.01	0.50	0.15	6.09	0.01	1.45	0.09	41.23
Benzo[g,h,i]perylene	BghiP	0.001	0.01	0.11	2.65	0.71	33.38	0.03	6.28	0.44	161
Σ16PAHs				7.62	84.85	18.65	755	1.87	336	18.81	11,353

Figure 3a shows high total PAHs values (ranging from 59.52 to 755 ng g⁻¹) in correspondence with urban towns such as Rome (755 ng g⁻¹), Naples (715 ng g⁻¹) and Palermo (303 ng g⁻¹). These areas are also the most densely populated and urbanised cities of the southern Italy. In a recent study, Albanese al. (2015a) highlighted that the principal source of the PAHs pollutants in the Neapolitan (Campania) soils is related to pyrogenic combustion (vehicular emission).

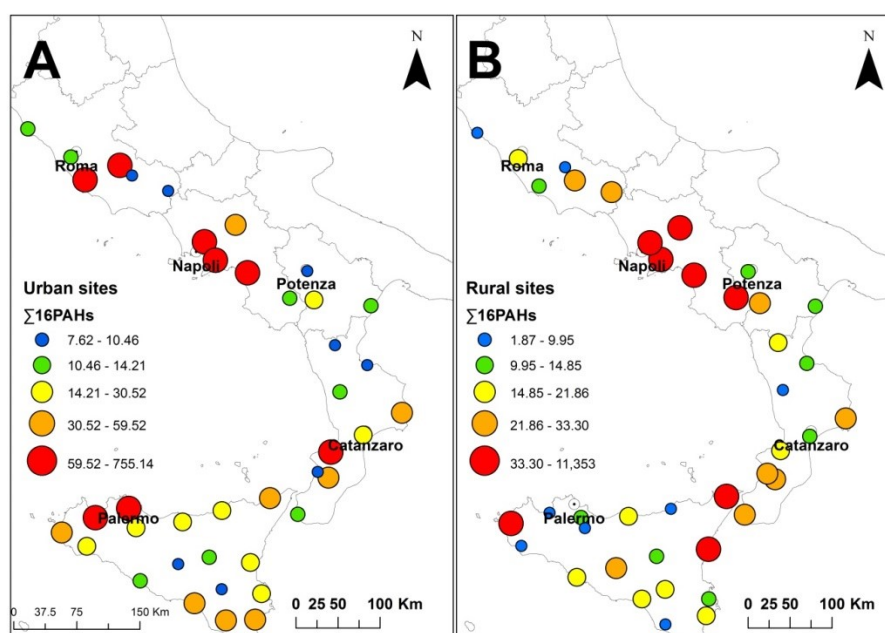


Fig. 3. Dots maps of the 16 PAHs concentration in urban (A) and rural (B) soils.

When considering the variation of total PAHs in rural areas (Fig. 3b), the highest values ranged from 33.30 to 11,353 ng g⁻¹ and are found mostly in the sampling sites of the Campania (Sarno Basin with 11,353 ng g⁻¹, Acerra with 917 ng g⁻¹ and Battipaglia with 276 ng g⁻¹), and the Sicily countryside (Acireale with 74.25 ng g⁻¹, Trapani with 56.56 ng g⁻¹ and Milazzo with 52.45 ng g⁻¹). When compared to other studies of PAHs in urban and rural soils worldwide, the ranges reported in our study present similar concentrations ranges, though with some noticeable differences (Table 3). Morillo et al. (2007) found total PAHs values ranging from 148 to 3410 ng g⁻¹ in the Turin (Italy) urban area; these values are four times greater than those found in urban areas of this survey study. The total 16 PAHs found in others European urban cities such in Seville (ranging from 89.5 to 4004 ng g⁻¹; Morillo et al. 2008), London (ranging from 400 to 67,000 ng g⁻¹; Vane et al. 2014), Glasgow (ranging from 48 to 51,822 ng g⁻¹; Morillo et al. 2007) and Moscow (208–3880 ng g⁻¹; Agapkina et al. 2007) are all higher than the values displayed in the urban areas of this study. Similarly, though at a larger scale, the ranges shown in the urban areas of Beijing (China) (Tang et al. 2005) and Delhi (India) (Bhupander et al. 2012) went from 219 to 27,825 ng g⁻¹ and from 81.6 to 45,017 ng g⁻¹, respectively. On the other hand, the rural areas of the southern Italy showed higher total concentration of PAHs compared to the values displayed in Hong Kong (China) (Zhang et al. 2006) and Delhi (Agarwal et al. 2009), ranging from 42.3 to 410 ng g⁻¹, and from 830 to 3880 ng g⁻¹, respectively. Many studies revealed that PAHs pollution sources are usually related to fuel combustion from traffic vehicle which is mostly occurring in urban areas. These studies are confirmed by the findings on some urban areas of the present study (e.g. high 16 PAHs in urban areas of Latium, Naples and Palermo).

However, this survey also highlighted some unexpected higher values outside of the urban areas, perhaps due to the presence of industrial activities such as incinerator plants, oils refineries as well as and illegal activities (e.g. wood and solid waste burning in the Acerra district and in the Naples wider metropolitan area). Even though such large industrial activities are actually forbidden in urban areas by the most recent Italian environmental legislation, their influence on the distribution of PAHs on rural areas remained, until now, largely unexplored.

These findings are confirmed by the Tukey's box and- whiskers plots (Fig. 4), which display how all 16 PAHs compounds showed higher concentration values in the rural sampling sites compared to the urban ones. On the other hand, Flo, Phe, Ant, Fluo, BaA, BbF, BkF, BaP, IcdP, DahA and BghiP, which are all HMWPAHs, displayed higher median concentration values in urban areas (Table 2)

Table 3. Total PAH concentrations (ng g⁻¹ dry weight) in the survey area topsoil compared to those found in other studies in the recent literature.

Study areas	Type of study area	Σ PAHs in soils (ng/g)	Authors
Southern Italy	Urban areas	7.62 - 755	this study
Southern Italy	Rural areas	1.87 - 11,353	this study
Turin, Italy	Urban areas	148 - 3,410	Morillo et al. (2007)
Hong Kong, China	Rural areas	42.3 - 410	Zhang et al. (2006)
Delhi, India	Urban areas	81.6 - 45,017	Bhupander et al. (2012)
Delhi, India	Rural areas	830 - 3,880	Agarwal et al. (2009)
Beijing, China	Urban areas	219 - 27,825	Tang et al. (2005)
Moscow, Russia	Urban areas	208 - 9,604	Agapkina et al. (2007)
Glasgow, UK	Urban areas	48 - 51,822	Morillo et al. (2007)
London, UK	Urban areas	400 - 67,000	Vane et al. (2014)
Seville, Spain	Urban areas	89.5 - 4,004	Morillo et al. (2008)

Since the median is a good, intuitive metric of centrality representing a 'typical' or 'middle' value (Reimann et al. 2008), it is reasonable to infer that HMWPAHs are therefore more likely to be related to those pollution sources that are occurring mostly in urban areas, such as vehicular emission and fuel combustion, whereas LMWPAHs could be more strongly related to biomass and oil combustion sources which are more likely to be occurring in rural areas (Chen et al. 2005; Aichner et al. 2007).

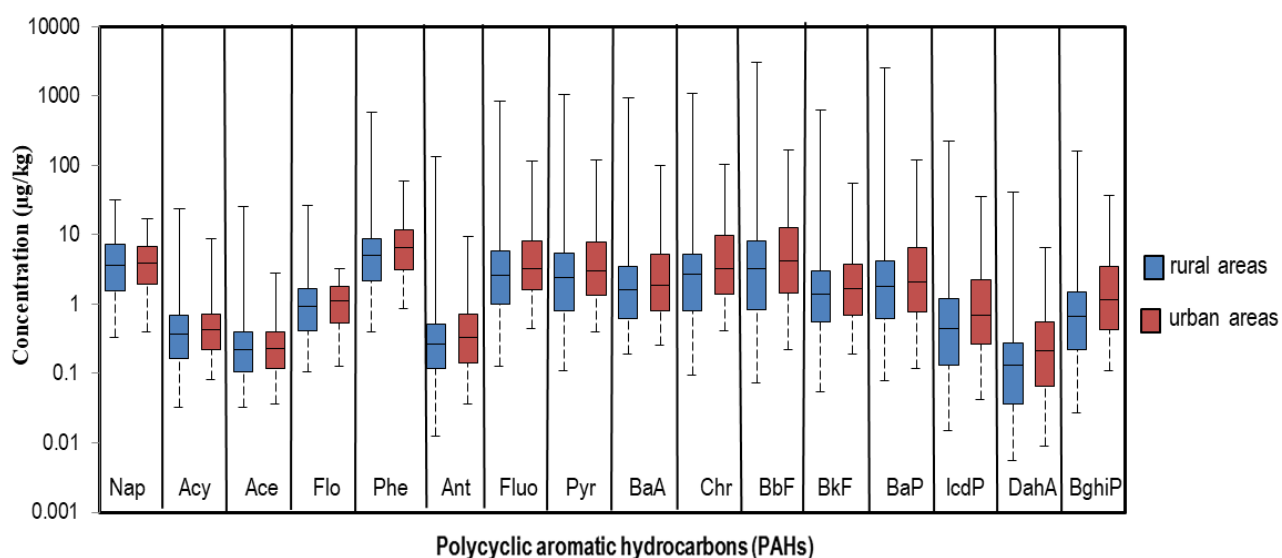


Fig. 4. Tukey's box-and-whiskers plot of individual PAH concentration (ng g⁻¹) in the survey area.

3.2. Diagnostic ratios and source apportionment of PAHs compounds

The LMWPAHs/HMWPAHs ratios (Fig. 5) were interpreted in terms of source apportionment. Soils from rural Sicilian areas displayed a higher mean value compared to those from urban areas, with ratios ranging from 0.12 to 2.48 (mean value of 1.14) compared to 0.08 to 1.35 (mean value of 0.56). These

values also indicate that the most likely sources of PAHs in Sicilian rural areas may be related to ‘petrogenic’ emissions. As mentioned earlier, in Sicily there are some of the most important Italian oil fields and refineries (e.g. Priolo, Gela, Ragusa and Milazzo—Bevilacqua and Braglia 2002). These industrial activities can give rise to ‘petrogenic’ emissions of PAHs and are found in suburban or nearby rural areas.

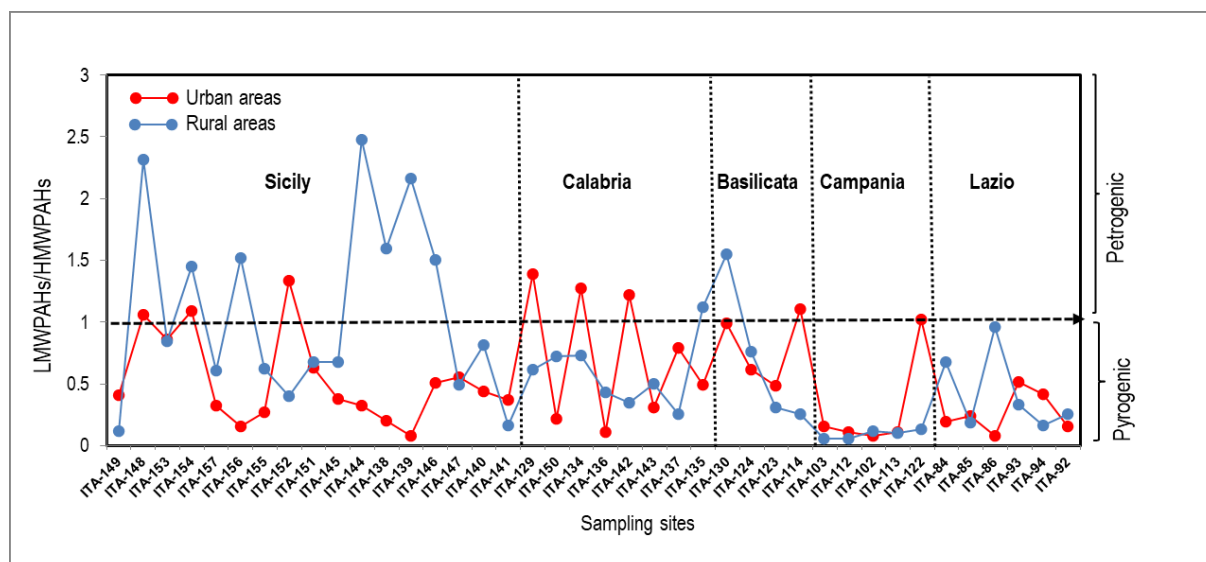


Fig. 5. LMPAHs/HMPAHs ratios scatter diagram of each sample site location. The different symbology (dots) reflects whether the sites were urban or rural nature.

In Calabria, the ratios are ranging from 0.11 to 1.37 (mean value 0.73) and from 0.25 to 1.11 (mean value 0.80) in urban and rural areas, respectively. In comparison, in Basilicata the ratios are ranging from 0.49 to 1.11 (mean value 0.59) and from 0.25 to 1.55 (mean value 0.71) in urban and rural areas, respectively.

From these values, it seems that Calabria displays some similar or slightly higher average values of LMWPAHs/HMWPAHs ratios compared to those in Basilicata. A potential explanation is that in Calabria there may be more sources of LMWPAHs (petrogenic, e.g. the large Crotone oil refineries) than in Basilicata. However, the highest measured ratios in Basilicata were found in the urban (ratio = 0.99) and rural (ratio = 1.55) areas of Viggiano municipality, in proximity of the most important Italian inland petroleum exploitation (Alta Val d’Agri oil field) (ISPRA 2014).

Campania and Latium presented lower ranges of LMWPAHs/HMWPAHs in urban areas, from 0.08 to 1.02 (mean value of 0.29) and from 0.08 to 0.52 (mean value of 0.27), respectively, and, from 0.08 to 0.14 (mean value of 0.10) and from 0.16 to 0.96 (mean value of 0.42) in rural areas, respectively. Given that most of the sites (but one) show a LMWPAHs/HMWPAHs ratio < 1, it seems that in these two areas the most likely sources of PAHs may be related to pyrogenic activities.

Individual molecular compound ratios were used to evaluate their potential sources (Hwang et al. 2003, Yunker et al. 2002). Fluo/(Fluo + Pyr) ratios ranged from 0.39 to 0.64, and 0.40 to 0.61 in urban and rural areas, respectively (Fig. 6a).

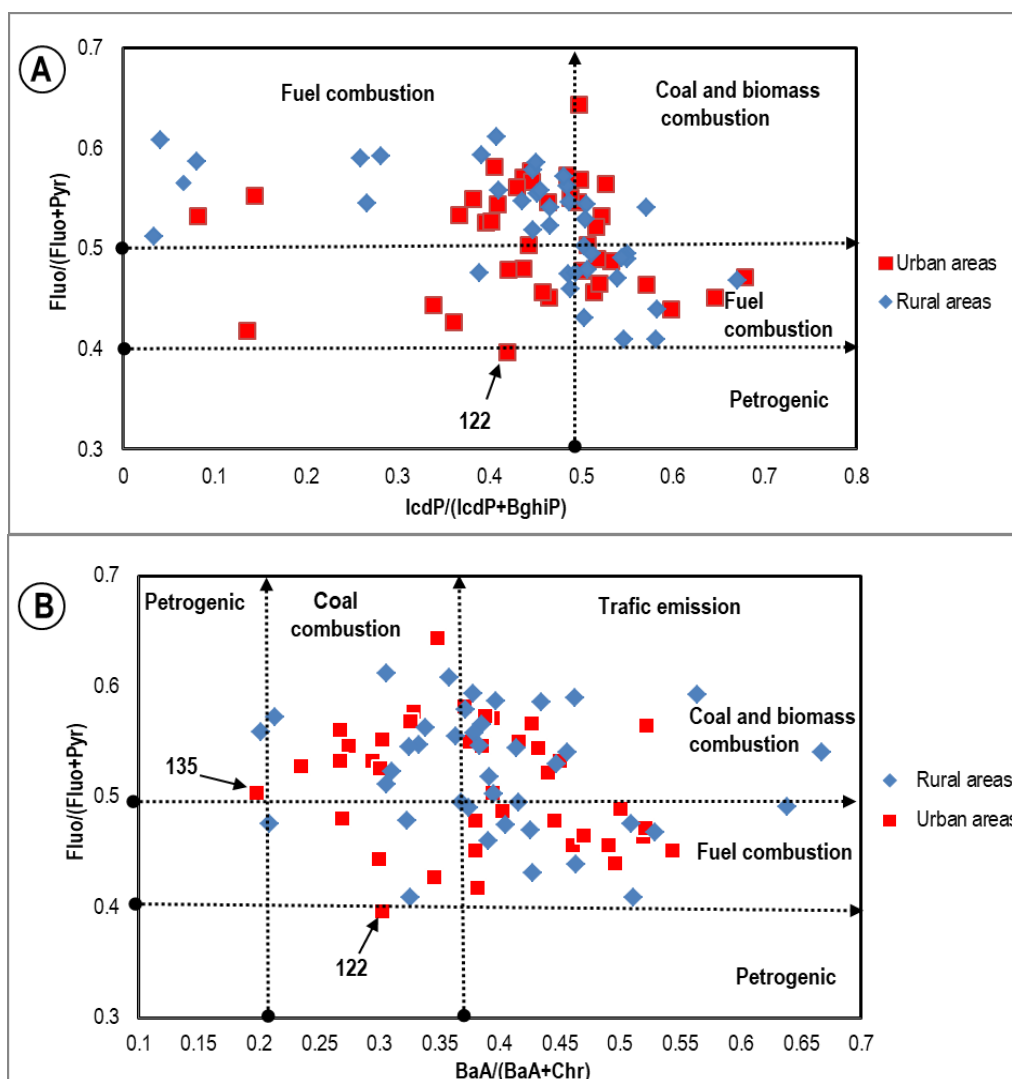


Fig. 6. Cross-plots for the isomeric ratios: (A) $\text{IcdP}/(\text{IcdP} + \text{BghiP})$ versus $\text{Fluo}/(\text{Fluo} + \text{Pyr})$ and (B) $\text{BaA}/(\text{BaA} + \text{Chr})$ ratio against $\text{Fluo}/(\text{Fluo} + \text{Pyr})$.

When considering sources characterisations ranges, it appeared that none of the sampling sites, neither urban nor rural areas, presented ratios characteristic of petrogenic sources (ratio < 0.4). Similarly, very limited sites displayed $\text{BaA}/(\text{BaA} + \text{Chr})$ ratio below 0.2. On the other hand, about 48.3 and 34.5% of the urban and rural sampling sites, respectively, presented a ratio that could be linked to a fuel combustion source (ratios from 0.4 to 0.5). This is also backed up by looking at two other diagnostic ratios: (1) the $\text{IcdP}/(\text{IcdP} + \text{BghiP})$ ratio, for which the results showed that the majority of the urban (58.2%) and rural (52.8%) areas fall in the range (<0.5) characterised by fuel combustion sources; (2) the $\text{BaA}/(\text{BaA} + \text{Chr})$ ratio (Fig. 6b), for which results indicated that the majority of the samples sites, in

urban and rural areas, presented ratios > 0.35 , corresponding to the same traffic combustion sources. Besides, Fluo/(Fluo + Pyr) ratios highlighted that 51.7 and 65.5% of urban and rural areas are displayed in the plot (ratios above 0.5) where sources of PAHs are more likely to be related to coal and biomass combustion. These measurements seem to suggest that the most likely sources of PAHs compounds in the studied areas may be related to fuel (vehicular and biomass) combustion (pyrogenic).

To summarise, by using different molecular diagnostic ratios, it can be highlighted that:

- LHMPAHs/HWMPAHs ratios \rightarrow dominant pyrogenic sources ($r < 0.1$)
- LHMPAHs/HWMPAHs ratios \rightarrow some petrogenic sources ($r > 1$ – Sicily).
- Fluo/(Fluo+Pyr) ratios \rightarrow no petrogenic source ($r < 0.4$)
- BaA/(BaA+Chr) ratios \rightarrow dominant pyrogenic sources ($0.4 < r < 0.5$)

3.3. Factor score maps for sources pattern

The total variance expressed by the 16 PAHs variables was 73.05% through three-factor models, F1, F2 and F3, accounting for 46.65, 15.05 and 11.35% of the variance, respectively (Table 4). Variables with loadings ≥ 0.50 were considered to describe the main composition of each factor. All variables hold communalities ≥ 0.5 (50% of variability), which means that they were all well correlated to one another. The associations of PAHs compounds for the three-factor models, sorted in descending loading values, were:

- F1: Fluo, Pyr, BaA, Chr, - (Nap, Acy, Flo)
- F2: BkF, BbF, BaP, - (Nap, Acy, Ace)
- F3: BghiP, DahA, IcdP, - (Ant)

The 16 PAHs variables were organised in two groups by using the log-transformed data and a varimax rotation in the factor analysis (Table 4), allowing to distinguish between positive and negative correlations within the three-factor models: G_1 = Fluo, Pyr, BaA, Chr, BbF, BkF, BaP, IcdP, DahA and BghiP (positive correlation) and G_2 = Nap, Acy, Flo, Nap, Acy, Ace and Ant (negative correlation).

These compound associations actually matched the main two groups of PAHs, where G_1 corresponds to the high molecular weight PAHs (HMWPAHs) and G_2 to the low molecular weight PAHs (LMWPAHs). In addition to this, each of the three-factor models displayed variables associations which can be used to further investigate and reveal the potential sources of the 16 PAHs contaminants in the survey area (Fig. 7).

Table 4. Varimax-rotated factor (three-factor model) of isometric logratio clr back-transformed variables for 80 topsoil samples from the survey area; bold entries: loading values over |0.50|.

Variables	Factors			Communalities
	F1	F2	F3	
Nap	-0.72	-0.58	-0.03	0.86
Acy	-0.51	-0.13	-0.42	0.51
Ace	-0.32	-0.66	-0.47	0.75
Flo	-0.55	-0.74	-0.27	0.93
Phe	-0.29	-0.73	-0.27	0.69
Ant	0.10	-0.07	-0.74	0.56
Fluo	0.92	-0.08	0.03	0.86
Pyr	0.92	0.27	-0.06	0.92
BaA	0.81	0.41	-0.10	0.83
Chr	0.71	0.30	0.28	0.67
BbF	0.30	0.79	0.21	0.75
BkF	-0.13	0.82	-0.05	0.69
BaP	0.44	0.76	-0.02	0.78
IcdP	0.24	0.38	0.69	0.68
DahA	-0.19	0.31	0.69	0.61
BghiP	0.25	-0.22	0.73	0.65
Eigenvalues	7.464	2.408	1.816	
Total variance in %	46.652	15.05	11.348	
Cum. of total variance (%)	46.652	61.703	73.051	

The F1 association (Fluo, Pyr, BaA, Chr, - (Nap, Acy, Flo)) accounted for the highest total variance (46.65%) with good adequacy (eigenvalues= 7.46 >1) between the factor and its variables.

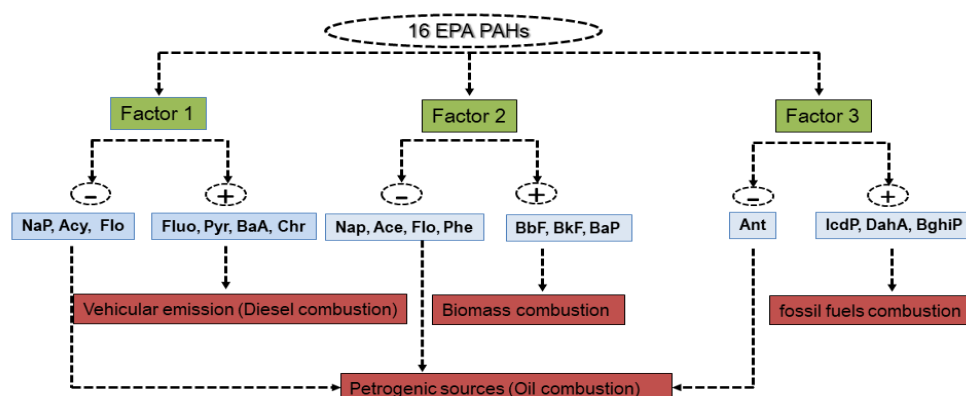


Fig. 7. Flowchart displaying the possible pollution sources of PAHs in the south of Italy throughout factor analysis. Each factor is characterised by antithetic variables association (marked by - and + symbols) which reveal the possible sources of PAHs in the survey area.

Factor scores distributions were processed in a GIS environment and interpolated to be displayed in maps to better visualise regional PAHs distribution (Lima et al. 2003; Thiombane et al. 2018). The F1 factor scores interpolated map (Fig. 8) presented the highest values (from 1.13 to 2.77) in and around the largest urban sites (Rome, Naples and Palermo). These areas are characterised by substantial vehicular traffic and emission (Spaziani et al. 2008). The lowest factor scores loadings (from - 2.52 to - 0.84) are mostly distributed in and around the rural sites in Basilicata (Val d'Agri oil field, Viggiano municipality), Calabria and Sicily (near the Priolo, Gela and Ragusa districts).

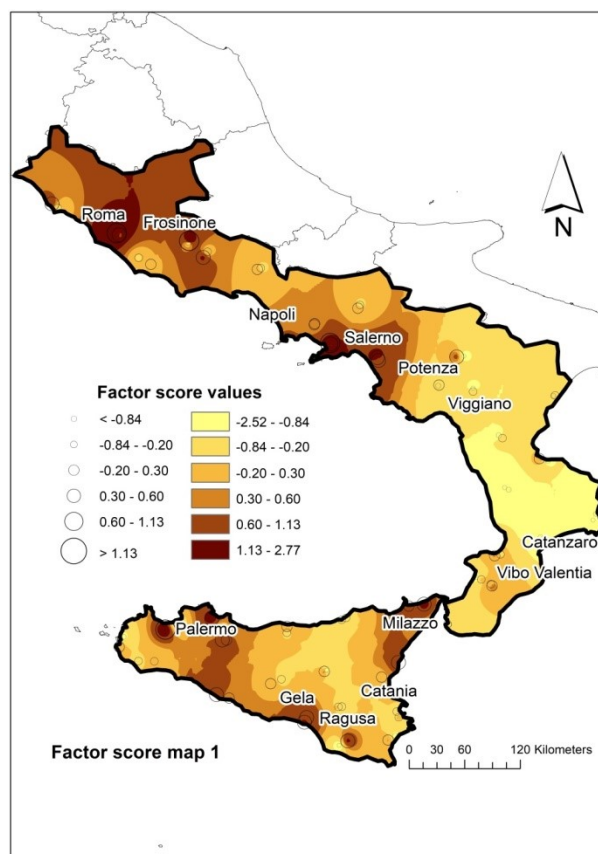


Fig. 8. Interpolated factor score map of the factor 1 (F1). Factor score values ranges are created by means of fractal concentration– area plot (C–A method).

Some of the compounds parts of this factor (Fluo, Pyr, BaA and Chr) are contaminants usually related to traffic and vehicle exhaust emissions (Sofowote et al. 2008; Li et al. 2012), which are obviously most likely to occur in urban areas. On the other hand, the other compounds parts of this factor (Nap, Acy and Flo) are instead usually indicative of spilled-oil-related products at low-temperature combustion (Yunker et al. 2002; Hwang et al. 2003; Bucheli et al. 2004) and can be associated with similar 'petrogenic' sources (e.g. oil combustion from petroleum exploitation industries) in the Basilicata and Sicily regions.

The F2 association (BbF, BkF, BaP, - (Nap, Acy, Ace)) expressed 15.05% of the total variance with an eigenvalue of 2.41. The F2 factor scores map (Fig. 9) presented the higher factor scores values (from 1.15 to 1.81) mostly distributed in and around the rural sites within the Campania region. In particular, the highest factor score values (from 1.81 to 2.23) were noted in the Acerra district. This municipality, which falls within the metropolitan/suburban area of Naples, is characterised by the presence of a large solid waste incinerator plant, but also by some illegal waste disposal and burning, and illegal practice of industrial toxic and solid urban waste dumping (Mazza et al. 2015; Marfe and Di Stefano 2016). At the same time, a number of industrial activities are also present in the nearby rural areas of the Campania region, where it is also common practice to use biomass resources for the combustion in heating systems. These results confirmed the findings of other studies which pointed out that BbF, BkF and BaP compounds were usually associated with biomass combustion (waste and wood combustion) in most rural areas in the Campania region (Arienzo et al. 2015; Albanese et al. 2015a).

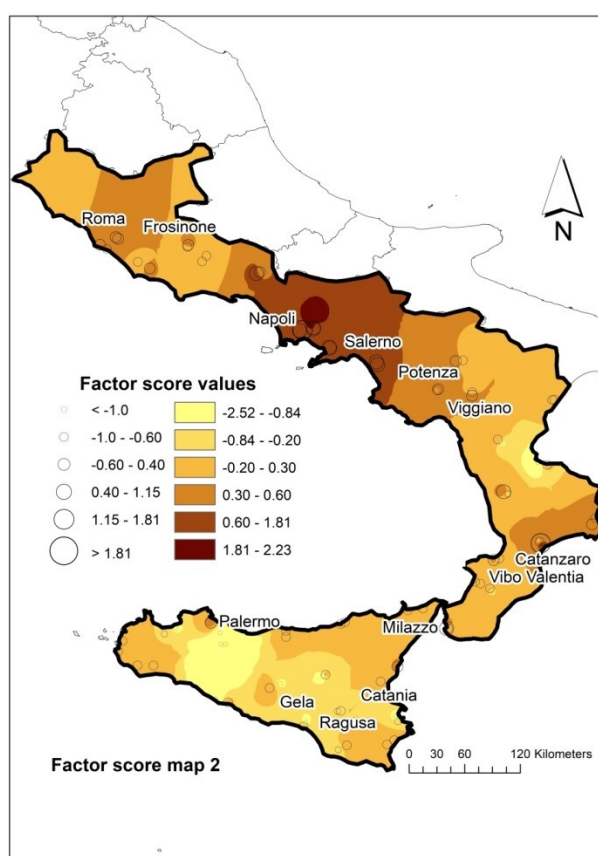


Fig.9. Interpolated factor score map of the factor 2 (F2). Highest factor scores are correlated with BbF, BkF and BaP compounds, and are displayed in the metropolitan area of Naples (Campania).

Similarly, Bixiong et al. (2006) suggested BbF and BkF as indicators of coal and wood combustions, whilst BaP has been used successfully as marker for biomass combustion (Simcik et al. 1999; Fang et al. 2004). The lower F2 factor scores (from - 2.55 to - 1.0) were found in the same areas where F1 (see

Fig. 8) displayed its lower factor score values. These areas were identified in Calabria (e.g. nearby Nucleo Industrial areas, Cosenza), and in areas nearby the oil field in Sicily (Priolo, Ragusa and Gela). These results support the observation that Nap, Acy and Ace compounds are usually related to oil combustion from petroleum exploitation industries (Masclet et al. 1987; Budzinski et al. 1997), which are present both in Basilicata and in Sicily.

Factor 3, contributing to 13.36% of the total variance, is dominated by positively correlated lcdP, DahA and BghiP and antithetic Ant. The factor score map (Fig. 10) presented high values (from 1.04 to 2.42) in and around most urban sites such as Naples, Vibo Valentia and Catania, and around the largest oil refineries in the study area (Milazzo and Val D'Agri). These finding are supported by the fact that lcdP, DahA and BghiP compounds are usually considered as markers of gasoline engine and crude oil combustion sources (Khalili et al. 1995; Larsen and Baker 2003).

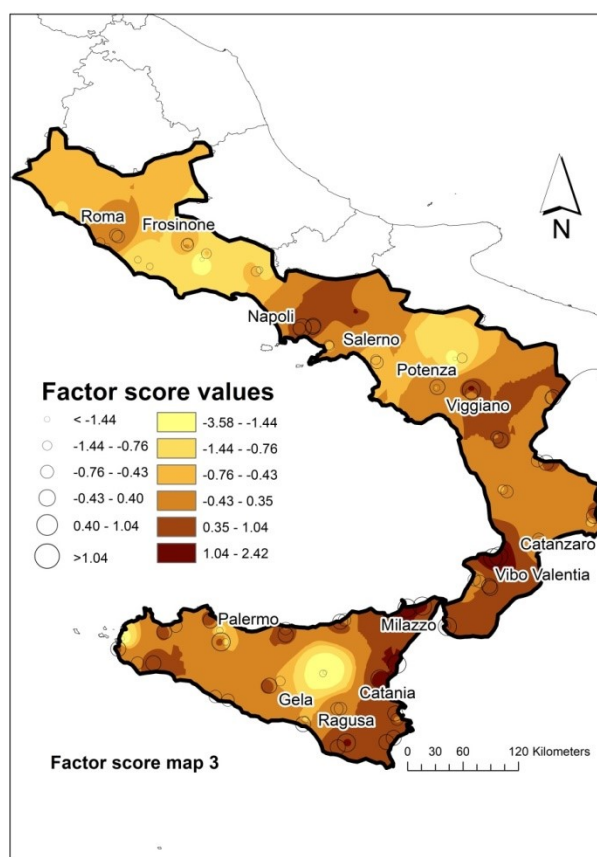


Fig. 10. Interpolated factor score map of the third factor (F3). High factor scores values corresponding to lcdP, DahA and BghiP variables are displayed in Naples, Vibo Valentia, Catania and around the largest oil refineries in the study area (Milazzo and Val D'Agri).

3.4. Potential soil toxicity

Using Eq. 1, potential toxicity levels of PAHs (in soils) were calculated and reported as TEQ_{BAP}. Results highlighted a significant variation of TEQ_{BAP} values in the five administrative regions studied (Fig. 11).

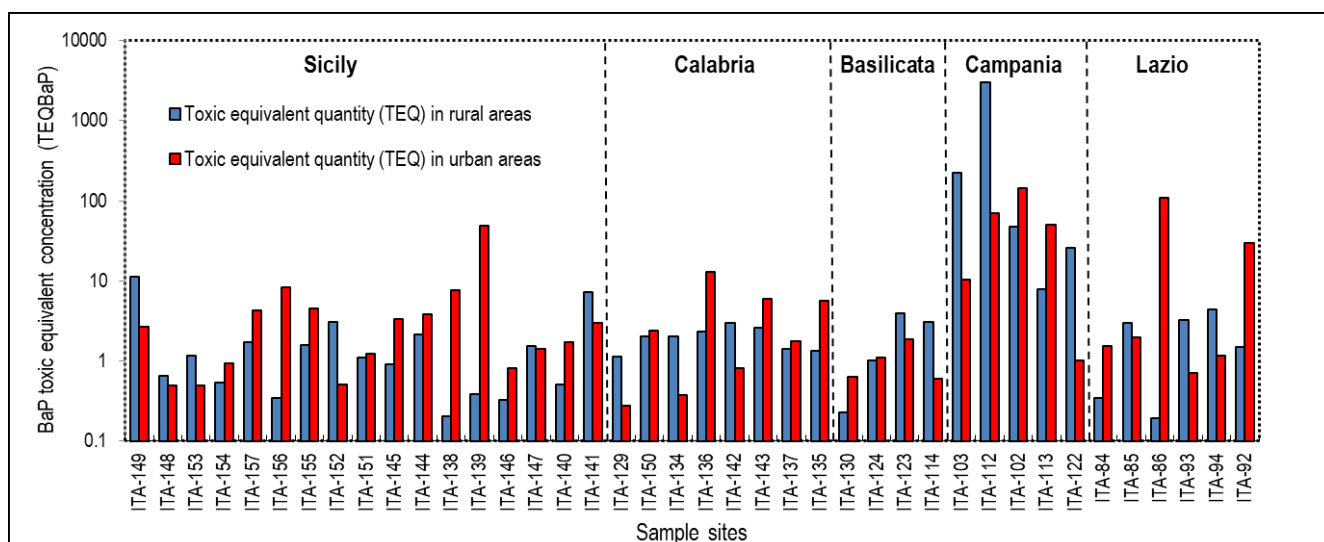


Fig. 11. Variation of the BaP toxicity equivalent quantity (TEQBaP) values in the survey area; logarithmic scale is applied on the Y axis.

TEQ_{BAP} values in Sicily ranged from 0.49 to 49.29 ng g⁻¹ and from 0.20 to 7.228 ng g⁻¹ in urban and rural areas with mean values of 5.56 and 2.04 ng g⁻¹, respectively, indicating that urban soils have higher TEQ_{BAP} toxicity levels compared to those in rural areas. Similarly, in Calabria and Latium average TEQ_{BAP} values were 3.40 and 23.87 ng g⁻¹ in urban areas, and 1.79 and 2.12 ng g⁻¹ in rural areas, respectively. The slightly higher mean value displayed by Latium (TEQ_{BAP} 23.87 ng g⁻¹) in urban areas can be associated with the prevalence of HWMPAHs in the TEQ computation. These compounds are usually related to vehicular emissions, where Rome (Latium region) is the Italian city with the highest motorisation rate (ISTAT 2012; ISPRA 2012).

Interestingly, Campania and Basilicata presented instead higher mean TEQ_{BAP} values in their rural areas (661 and 2.71 ng g⁻¹, respectively), whilst the mean TEQ_{BAP} values for their urban areas were 54.93 and 1.19 ng g⁻¹, respectively.

Overall, Campania presented the highest toxicity average values in both its urban and rural areas. The values found in the present study are significant when compared to similar works published worldwide. In urban areas, variability is greater and depends on size of the city and urban fabric, For example, TEQ_{BAP} values found in Gwangju City (Korea) were 13.23 ng g⁻¹ (Islam et al. 2017), which are comparable to those found in urban areas of Tarragona (Spain) with 64.0 ng g⁻¹ (Nadal et al. 2007) and the present study, whilst those found in urban soils of a megacity such as Delhi (India) were an order of magnitude higher (218 ng g⁻¹ - Agarwal et al. 2009), (Table 5).

Table 5. Means TEQBaP values in the five studied regions compared to other selected survey studies.

Regions	Type of area	TEQ _{BaP} (ng/g)	References
Sicily	Urban	5.56	this study
	Rural	2.04	this study
Calabria	Urban	3.40	this study
	Rural	1.79	this study
Basilicata	Urban	1.19	this study
	Rural	2.71	this study
Campania	Urban	54.93	this study
	Rural	661	this study
Latium	Urban	23.87	this study
	Rural	2.12	this study
Gwangju City (Korea)	Urban	14.30	Islam et al. (2017)
Tarragonain (Spain)	Urban	64	Nadal et al. (2007)
Delhi (India)	Urban	218	Agarwal et al. (2009)
Rural soils in Norway	Rural	14.3	Nam et al. (2008)
Agriculture soils in Poland	Agriculture soil	11.9	Maliszewska-Kordybach et al. (2009)
Rural soils in the UK	Rural	83.5	Nam et al. (2008)

In rural soils studies, values found in the literature vary from 11.20 ng g⁻¹ in agricultural soils in Poland (Maliszewska-Kordybach et al. 2009), 14.30 ng g⁻¹ in some soils of Norway (Nam et al. 2008), to values of 83.12 ng g⁻¹ found in rural soils in the UK (Nam et al. 2008) (Table 5). Several studies (Means et al. 1980; Agapkina et al. 2007; Bhupander et al. 2012) revealed that the occurrence of individual PAHs in soils depends on the land use and the settlement of main PAHs pollution sources. In fact, biomass burning is considered the major source of PAHs in rural soils, and urban soils are likely polluted by release of PAHs from vehicular emission around the heavy traffic roads (Khalili et al. 1995; Morillo et al. 2007; Albanese et al. 2015a). This study unveiled that urban areas of Sicily, Calabria and Latium displayed higher TEQ_{BaP} values than in their rural areas, but rural soils of Basilicata and Campania revealed the contrary. Among all the findings, it is perhaps arguable that the most striking of all is that related to the metropolitan and rural areas around Naples, characterised by high toxicity levels. These results are partially confirmed by a more detailed study carried out by Albanese et al. (2015a) and by a larger study (known as Campania Trasparente - still in progress), with sampling of soils and air matrices covering the entire regional territory (Qu et al. 2018).

4. Conclusions

This study carried out a regional survey of urban and rural topsoils (80 samples) in five administrative regions of the south of Italy, in an attempt to shed light to the main potential sources and patterns of pollution for 16 PAHs compounds (US EPA priority compounds).

Measuring molecular PAHs diagnostic ratios allowed to identify and clearly show the main areas of concern, as well as giving an indication of the most likely sources for the PAHs compounds in the various regions. Results strongly pointed to the direction of pyrogenic sources (e.g. vehicular emission, fuels and biomass combustions) for some areas (urban areas of Latium and rural areas of Campania regions). In particular, these areas were found to have high BAP concentration levels, which in turn indicated high levels of toxicity equivalent quantity (TEQ_{BAP}). The highest level of TEQ (661 ng g^{-1}) was found in metropolitan and rural areas of the Campania region, which could be related to the presence of a large solid waste incinerator plant, illegal waste disposal and burning, and illegal practice of industrial toxic and solid urban waste dumping (both known in the area).

Results from this study can represent a fundamental step to understand the distribution, sources and toxicity levels of PAHs in the soils of these regions, giving an impetus to follow up with more detailed surveys. Given the carcinogenic and mutagenic properties of these contaminants, it is envisaged that the findings from this study will help initiate an assessment of human health risks from to PAH exposures in rural and urban areas in this Mediterranean region.

Section 3.2

Status, sources and contamination levels of organochlorine pesticide residues in urban and agricultural areas: a preliminary review in central-southern Italian soils

This section has been published in:
Journal of Environmental Science and Pollution Research
Volume 25, Issue 26, September 2018, pages 361–382

Status, sources and contamination levels of organochlorine pesticide residues in urban and agricultural areas: a preliminary review in central–southern Italian soils

Abstract

Organochlorine pesticides (OCPs) are synthetic chemicals commonly used in agricultural activities to kill pests and are persistent organic pollutants (POPs). They can be detected in different environmental media, but soil is considered an important reservoir due to its retention capacity. Many different types of OCPs exist, which can have different origins and pathways in the environment. It is therefore important to study their distribution and behaviour in the environment, starting to build a picture of the potential human health risk in different contexts. This study aimed at investigating the regional distribution, possible sources and contamination levels of 24 OCP compounds in urban and rural soils from central and southern Italy. One hundred and forty-eight topsoil samples (0–20 cm top layer) from 78 urban and 70 rural areas in 11 administrative regions were collected and analysed by gas chromatography–electron capture detector (GC–ECD). Total OCP residues in soils ranged from nd (no detected) to 1043 ng/g with a mean of 29.91 ng/g and from nd to 1914 ng/g with a mean of 60.16 ng/g in urban and rural area, respectively. Endosulfan was the prevailing OCP in urban areas, followed by DDTs, Drins, Methoxychlor, HCHs, Chlordane-related compounds and HCB. In rural areas, the order of concentrations was Drins > DDTs > Methoxychlor > Endosulfans > HCHs > Chlordanes > HCB. Diagnostic ratios and robust multivariate analyses revealed that DDT in soils could be related to historical application, whilst (illegal) use of technical DDT or dicofol may still occur in some urban areas. HCH residues could be related to both historical use and recent application, whilst there was evidence that modest (yet significant) application of commercial technical HCH may still be happening in urban areas. Drins and Chlordane compounds appeared to be mostly related to historical application, whilst Endosulfan presented a complex mix of results, indicating mainly historical origin in rural areas as well as potential recent applications on urban areas. Contamination levels were quantified by Soil Quality Index (SoQI), identifying high levels in rural areas of Campania and Apulia, possibly due to the intensive nature of some agricultural practices in those regions (e.g., vineyards and olive plantations). The results from this study (which is in progress in the remaining regions of Italy) will provide an invaluable baseline for OCP distribution in Italy and a powerful argument for follow-up studies in contaminated areas. It is also hoped that similar studies will eventually constitute enough evidence to push towards an institutional response for more adequate regulation as well as a full ratification of the Stockholm Convention.

1. Introduction

The Stockholm Convention (2005) banned the use of Persistent Organic Pollutants (POPs), with the aim of protecting human health and the environment. The initial list prepared in 2003 included Aldrin, Dieldrin, Endrin, Chlordane, Heptachlor, Hexachlorobenzene (HCB) and Dichlorodiphenyltrichloroethane (DDT).

This list was then expanded with other potential POPs in 2011 (Stockholm Convention on Persistent Organic Pollutants 2011): Hexachlorocyclohexane (HCH, including Lindane), Methoxychlor, and Endosulfan (Stockholm Convention 2005, 2011). These organic pollutants are considered long-range transport compounds based on their ubiquity, persistence and bioaccumulation potential in different environmental media (Weinberg 1998; Szeto and Price 1991; Fang et al. 2017), as well as their high toxicity to humans and non-target organisms (WHO 2003; Nizzetto et al. 2006; Moeckel et al. 2008; Kim et al. 2017). They sink in different environmental matrices, such as air, water and soils, and further accumulate in the food chain (Prapamontol and Stevenson 1991; Suchan et al. 2004; Qu et al. 2016). Soil continues to be a potential medium of exposure of OCPs, and its biofilms and physico-chemical properties may influence fate and behaviours of OCP metabolites through different degradation phenomena (Weinberg 1998; WHO 2003).

Since agricultural practices are a very important economic resource for Italy, this makes it the third OCP user among European Union countries (Eurostat 2014). In Italy, OCPs are used in most agricultural activities, in forestry as well as ornamental plants in urban garden preservation against insects, fungal or animal pests. It is well-known that Italy is the only European Union (EU) country that has not ratified the Stockholm Conventions, though the production and use of Aldrin, Chlordane, Dieldrin, Endrin, DDT, Heptachlor, HCB and HCH in its territory have been strictly restricted in harmony with several other regulatory schemes via the Rotterdam Convention in 1998, the European Directive in 2000 (Persistent organic pollutants amending Directive 79/117/EEC) and the United Nations Economic Commission for Europe POPs Protocol (UNECE 2010). Moreover, Italian environmental law (D. Lgs. 152/2006) established guideline threshold values that regulate the mitigation of OCPs in soils. This regulation guided a recent evaluation of the levels (for DDT) and residues (for HCH) in the Campania plain (Arienzo et al. 2015; Qu et al. 2016) and in agricultural soils in the province of Latina (Latium) (Donnarumma et al. 2009). However, this legislation does not involve OCPs such as Endosulfan and Methoxychlor: these compounds have been associated to both environmental and human health risks due to concerns that they are carcinogen, teratogen and male reproductive toxicants (PANNA 2008; US EPA 2007; Silva and Carr 2009; Jayaraj et al. 2016). Whilst recent studies have started to investigate and define the level of OCPs in Italian soils (e.g., Donnarumma et al. 2009; Arienzo et al. 2015; Qu et al. 2016, 2017), there has been no systematic attempt to evaluate their wider distribution and variations

across rural and urban areas in Italy. The aim of this study was to begin to establish a regional (and eventually national) baseline based on a large survey carried out in 11 regions of central and southern Italy. The main objectives of this study were as follows:

- (1) to identify the regional distribution of OCP compounds in Italian soils;
- (2) to evaluate their potential sources by using OCP diagnostic ratios as well as robust compositional biplot and factor analysis; and
- (3) to quantify OCP contamination levels by using Soil Quality index (SoQI) in urban and rural soils.

This study is important because it will represent a fundamental stepping stone to build a long-overdue national picture of OCP status in Italy. It is envisaged that the results of this study should trigger more detailed surveys in contaminated areas as well as ad hoc risk-based studies, which in the long term will constitute a strong-enough argument to cause an adequate institutional response by the Italian regulating authorities.

2. Materials and methods

2.1. Study area

The survey area included 4 administrative regions (Latium, Marches, Tuscany, and Umbria) from central and 7 (Abruzzo, Apulia, Basilicata, Calabria, Campania, Molise, and Sicily) from southern Italy (Fig. 1).

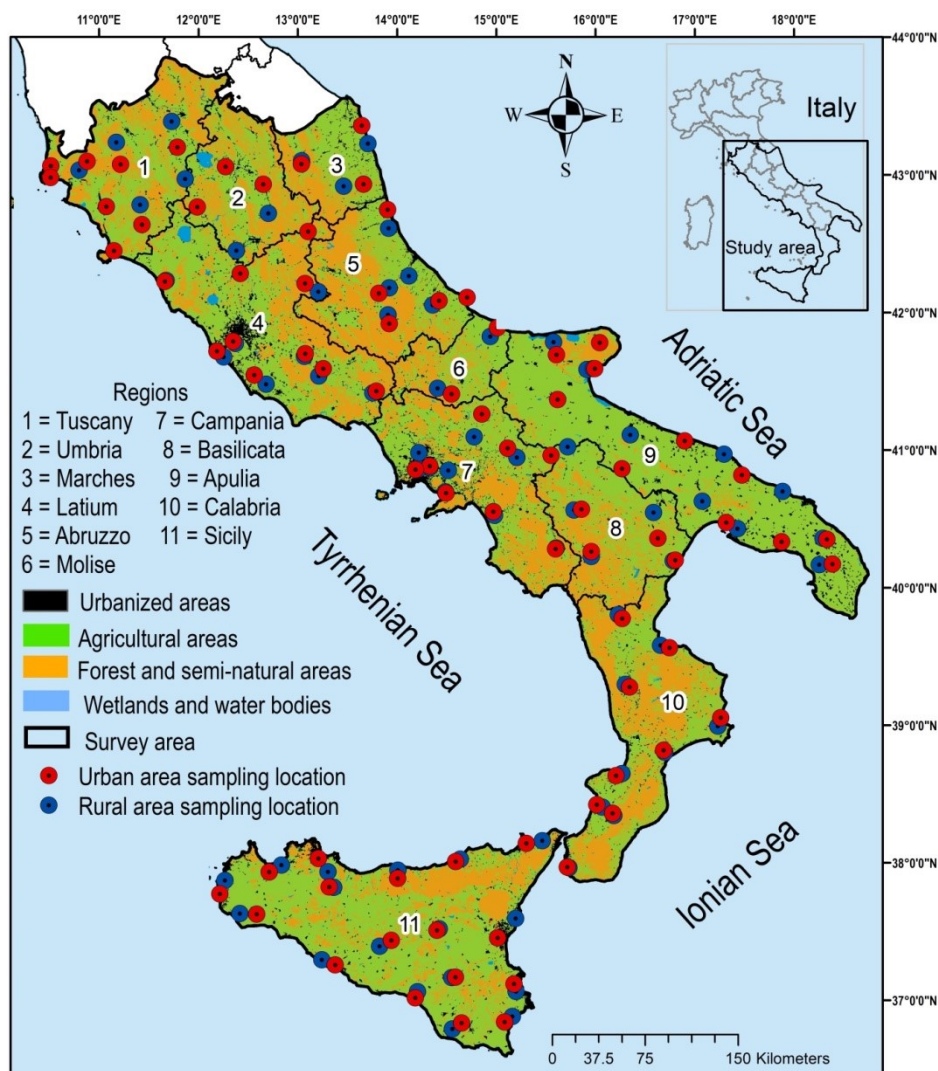


Fig. 1. Land use/land cover of the study area (simplified from Corine Land Cover 2012). Urban (red dots) and rural (blue dots) sampling site locations are displayed.

The total survey area (considering administrative regional boundaries) extended to 157,716 km² with 31.26 million of inhabitants, mostly grouped in main urban areas (ISTAT, 2016). Most of the land is used for agricultural and forestry activities. Agriculture occupies one-fourth of the land available, which includes cultivation of hilly areas where agriculture results in modifying the natural landscape and resources through terracing, irrigation, and soil management (Corona et al., 2012; ISTAT, 2013; ISPRA, 2014a). Favourable meteorological conditions, dominated by a Mediterranean climate, allow intensive agriculture activities such as vineyards and olive plantations - mostly in hilly areas – as well as greenhouse production (tomatoes, potatoes, aubergines, peppers, peas, and citrus fruits) in coastal areas in Campania, Apulia, and Sicily (Costantini and Dazzi, 2013). The predominant crops in inland territory are seasonal ones like wheat, maize, potatoes, rice, and sugar beet. Most of forestry lands are composed of broad-leaved trees, with conifers and chestnut making up about one-fifth of the total (ISTAT, 2013).

Large urban areas such as Rome (Latium), Naples (Campania), Bari (Apulia) and Palermo (Sicily), are densely populated and surrounded by metropolitan areas where both industrial activities, manufactories and intensive agriculture occur (ISTAT, 2016).

2.2. Soil sampling procedure and preparation

The sampling campaign took place from early April to end of September 2016, with the aim to select the most representative topsoil samples in urban and rural areas throughout 11 regions (Latium, Marche, Tuscany, Umbria, Abruzzo, Apulia, Basilicata, Calabria, Campania, Molise, and Sicily) from the centre to southern Italy. In each region, the main urban areas and the nearest rural areas where most of the land is devoted to agricultural activities, were selected. Site selection was carried out by interpreting, using Geographical Information Systems (ArcGIS, 2012), information on land use/land cover of the study area (ISPRA, 2014b; Corine land cover, 2012) together with satellite imagery (Google Earth® professional, 2016). A total of 148 soil samples were collected with a nominal density of 2 samples/ 2500 km² (in urban and rural areas) (Fig. 1). Samples have been collected from public gardens in urban areas, and from agricultural land (farmlands/cropland) in rural areas. All the samples were collected using a stainless steel scoop, kept in labelled glass bottles and directly stored in ice boxes to minimize the losses caused by volatilization and initial degradation of the organic compounds. Each topsoil sample (from 0-20 cm) was made by homogenizing 5 subsamples at the corners and the centre of a 100m² square, collecting approximately 1.5 kg in total. The sampling procedure followed the Geochemical Mapping of Agricultural and Grazing Land Soil (GEMAS) sampling procedure described by Reimann et al. 2014). Soil samples were homogenized and sieved using a <2mm mesh sieve after removing stones, detritus and residual roots. Finally, composite samples were stored at -4 °C in the environmental geochemistry laboratory of the University of Naples Federico II (Italy) until instrumental analysis. Geographical coordinates were recorded by geospatial positioning systems (WGS84, GPS) at each sample site.

2.3. Extraction procedure and analysis OCPs

Analyses were carried out by an Agilent 7890A gas chromatograph with a ⁶³Ni electron capture detector (GC-ECD) equipped with a DB-5 capillary column (30.0 m length, 0.32 mm diameter, 0.25 mm film thickness), in the Key Laboratory of Biogeology and Environmental Geology of Ministry of Education at the University of Geosciences in Wuhan, China (Yang et al. 2008; Qu et al., 2016).

Gas chromatography-mass spectrometry (GC-MS) and gas chromatography-electron capture detector (GC-ECD) are the most common and appropriate systems to investigate organic contaminants in different environmental media. Many authors (Aramendia et al., 2007; Alves et al., 2012) showed the high sensitivity of GC-ECD for organophosphorus and organochlorine pesticides. In this study, the

rationale of working with GC-ECD analyser was based on the excellent sensitivity and satisfactory quantification limits, allowing the identification and quantification of pesticides at low levels. A 10 g of dried soil samples were spiked with 20 ng of 2,4,5,6-tetrachloro-m-xylene (TCmX) and decachlorobiphenyl (PCB209) as recovery surrogates and were Soxhlet-extracted with dichloromethane for 24 h. Activated copper granules were added to the collection flask to remove elemental sulphur. The extraction of OCPs was concentrated and solvent-exchanged to n-hexane and further reduced to 2–3 mL by rotary evaporation. The alumina/silica (1:2) gel column (450°C muffle drying for 4 h, both deactivated with three percent water) was used to purify the extract and OCPs were eluted with 30 mL of dichloromethane/hexane (2/3). Then the eluate was concentrated to 0.2 mL under a gentle nitrogen stream and a known quantity of penta-chloronitrobenzene (PCNB) was added as an internal standard prior to gas chromatography–electron (GC–ECD) analysis.

Nitrogen was used as carrier gas at 2.5 mL/min under constant-flow mode. Injector and detector temperatures were maintained at 290°C and 300°C, respectively. The oven temperature started from 100°C (with an equilibration time of 1 min), and rose to 200°C at a rate of 4°C/min, then to 230°C at 2°C/min, and finally reached 280°C at 8°C/min, and was held for 15 min. 2 µL of each sample was injected into the GC-µECD system for the analysis. Concentration of the individual target OCPs were identified by comparison of their retention times (previously confirmed with GC/MS) and quantified using an internal standard. The gas chromatograph (GC-MS) parameters of the Agilent 6890GC-5975MSD system were the same as those of the Agilent 6890 GC equipped with 63Ni micro-electron capture detector (GC-µECD). The mass spectrometer (MS) was operated in electron impact ionization mode with electron energy of 70 eV. The ion source, quadrupole and transfer line temperatures were held at 230, 150 and 280° C, respectively. Target compounds were monitored in selected ion monitoring (SIM) mode.

Procedure types used for quality assurance and quality/control (QA/QC) were as follows: method blank control (procedural blank samples), parallel sample control (duplicate samples), solvent blank control, and basic matter control (US EPA, 2000). The spiked samples containing internal standard compounds were analysed simultaneously with soil samples. A procedural blank and a replicate sample were run with every set of 12 samples analyzed to check for contamination from solvents and glassware. The limits of detection (LODs) were based on 3:1 S/N ratio. TCmX and PCB 209 were spiked as surrogate standards to judge procedural performance. The surrogate recoveries for TCmX and PCB 209 were $77.8 \pm 19.0 \%$ and $89.3 \pm 20.3 \%$, respectively. The relative standard deviation (RSD) was less than 10%. All OCPs concentrations were expressed on an air-dried weight basis.

2.4. Geostatistical and multivariate analysis

OCPs associations and possible sources were identified by univariate and multivariate statistical analyses as well as diagnostic ratios, compositional biplot and robust factor analysis. Compositional biplot and robust factor analysis allowed to minimize and/or eliminate the presence of outliers and spurious correlation (Pawlowsky-Glahn and Buccianti, 2011; Filzmoser et al., 2012). DDT and HCH compounds were chosen for the multivariate computation both for their high toxicity levels, and for their proven predominance in Italian soils and air (e.g., Estellano et al., 2012; Pozo et al., 2016; Qu et al., 2016). Biplot statistical analysis (Gabriel, 1971) was used to display both samples and variables of the data matrix in terms of the resulting scores and loading (Pison et al., 2003; Otero et al., 2005). For a full description of compositional biplot, several examples are available in the literature (e.g., Maronna et al., 2006; Filzmoser et al., 2008, 2009; Hron et al., 2010; Thiombane et al., 2018). Factor analysis (FA) was used to explain the correlation structure of the variables through a reduced number of factors (Reimann et al., 2002). This has been successfully employed to evaluate the potential origins of the compounds in relation to their main hypothetical sources (Reimann et al., 2002; Jiang et al., 2009). Isometric logratio transformation (ilr) was applied on raw data prior to multivariate analysis (Filzmoser et al., 2009). R-mode factor analysis was also performed, and the different factors obtained studied and interpreted in accordance with their presumed sources (Reimann et al., 2002; Albanese et al., 2007).

Two main open-source R packages for statistical software were used: “Compositions” (Van Den Boogaart et al., 2011) and “Robcompositions” (Templ et al., 2011). OCPs concentrations and factor score values were mapped for image-patterns recognition using GeoDAS (Cheng et al., 2001) and ArcGIS (ESRI, 2012) software. GeoDAS™ was used to produce dots and interpolated geochemical maps using the multifractal inverse distance weighted (MIDW) algorithm (Cheng et al., 1994; Lima et al., 2003). The concentration–area (C–A) fractal method was applied to classify OCPs concentration and factor score ranges in interpolated images.

2.5. Assessment of contamination level

Assessment of contaminated sites is a preliminary requirement to reveal potential impact of OCPs pesticides on public and ecosystem health (USEPA, 1991; CCME, 1992; Doe, 1995; APAT, 2008; DEFRA, 2011). The “Soil Quality Index” (SoQI) elaborated by the Canadian Soil Quality Guidelines for Protection of Environment and Human Health Agency (CCME, 2007) was implemented to define, classify and prioritize contamination level for each region. Advantages of the SoQI include that it is a robust computation based on three factors for its calculations, namely: 1) scope (% of contaminants that do not meet their respective guidelines), 2) frequency (% of individual tests of contaminants that do not meet their respective guidelines), and 3) amplitude (the amount by which the contaminants do not meet

their respective guidelines) and it is relatively simple to use. The SoQI was computed using thresholds values for residential areas established by Italian environmental law (D. Lgs. 152/2006) (Table 1) as reference guidelines.

Table 1. Organochlorine pesticide guideline threshold values in soils, fixed by the Italian environmental law (D. Lgs. 152/2006) in residential areas (and/or park areas) and industrial (or/or commercial) sites.

SoQI index provides a quantitative index based on the amalgamation of the three factors (F_1 , F_2 and F_3):

$$F1 = \frac{\sum fx}{\sum Cx} \times 100 \quad (1)$$

$F1$ (scope) represents the percentage of contaminants that do not meet their respective guideline values, where fx is the number of failed contaminants, and Cx is the total number of contaminants

$$F2 = \frac{\sum ftx}{\sum tx} \times 100 \quad (2)$$

$F2$ (frequency) corresponds to the percentage of individual tests that do not meet their respective

	Residential or recreation or park areas sites (ng/g)	Industrial or commercial sites (ng/g)
Aldrin	10	100
α -HCH	10	100
β -HCH	10	500
γ -HCH or Lindane	10	500
δ -HCH	10	100
Chlordane	10	100
DDT, DDE, DDD	10	100
Dieldrin	10	100
Endrin	10	2000

guidelines values, ftx represents the number of failed tests and tx symbolizes the number of tests.

$$Ex_i = \frac{Zt_i}{Gv_i} - 1 \quad (3)$$

Ex_i or Excursion is the magnitude by which the contaminant is over/below the respective guideline value.

This is calculated as a ratio of the failed test value (Zt_i) and its respective guideline value (Gv_i)

$$Ase = \frac{\sum_{i=1}^n Ex_i}{\sum ftx} \quad (4)$$

The average amount by which individual tests are out of compliance corresponds to Ase .

$$F3 = \frac{Ase}{0.01Ase + 0.01} \quad (5)$$

$F3$ or amplitude represents the amount by which failed test values do not meet their guidelines.

$$\text{SoQI} = 100 - \frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732} \quad (6)$$

And finally, SoQI is calculated by taking the square root of the sum of squared factors divided by 1.732 and extracting it from 100. The 1.732 normalizes the SoQI to a range between 0 and 100. The proposed classes are: very low contamination (90-100), low contamination (70-90), medium contamination (50-70), high (30-50) and very high contamination (0-30).

3. Results and discussion

3.1. Residues and pollution sources of OCPs

3.1.1. OCPs

Total OCPs residues in soils ranged from “no detected” (nd) to 1043.98 ng/g with a mean of 29.91 ng/g, and from nd to 1914.1 ng/g with a mean of 60.16 ng/g in urban and rural area, respectively (Table 2).

The coefficient of variation (CV) ranged from 0.27 to 8.72, and from 1.87 to 6.47 in urban and rural areas, respectively, reflecting a significant spatial variation.

Table 2. Descriptive statistics of the 24 OCP compounds (ng/g) in 148 topsoil samples from urban and rural areas of central and southern Italy; min, max and CV indicate the minimum, maximum and coefficient of variation of the dataset, respectively.

Compounds (ng/g)	DL	Urban				Rural			
		Min.	mean	Max.	CV	Min.	Mean	Max.	CV
α -HCH	0.011	n.d	0.22	4.43	2.40	n.d	0.57	19.21	4.15
β -HCH	0.006	n.d	1.10	5.50	1.36	n.d	1.47	20.38	1.87
γ -HCH	0.011	n.d	0.39	14.19	4.19	n.d	0.65	11.29	2.73
δ -HCH	0.01	n.d	0.11	2.72	3.06	n.d	0.36	18.18	6.05
HCHs	-	n.d	1.82	25.08	1.78	n.d	3.05	47.27	2.30
p,p'-DDT	0.025	n.d	1.81	16.99	1.73	n.d	8.90	418.46	5.68
o,p'-DDT	0.02	n.d	0.34	5.04	2.24	n.d	1.68	48.27	4.80
p,p'-DDE	0.019	n.d	1.81	38.59	3.20	n.d	5.40	139.93	4.19
o,p'-DDE	0.021	n.d	0.33	7.56	2.99	n.d	0.27	4.34	2.26
p,p'-DDD	0.006	n.d	0.28	3.05	2.19	n.d	1.11	36.22	4.24
o,p'-DDD	0.025	n.d	0.34	5.04	2.24	n.d	1.68	48.27	4.80
DDTs	-	n.d	5.26	56.98	1.90	n.d	18.01	632.95	4.59
cis-Chlordane	0.018	n.d	0.08	1.77	2.93	n.d	0.11	3.40	4.07
trans-Chlordane	0.021	n.d	0.20	7.71	4.47	n.d	0.10	1.97	3.33
Heptachlor	0.021	n.d	0.15	3.72	3.41	n.d	0.07	1.20	2.33
Heptachlor-epoxide	0.014	n.d	0.62	10.95	2.64	n.d	0.57	13.73	3.43
Aldrin	0.046	n.d	0.25	2.69	1.75	n.d	0.59	13.37	3.27
Dieldrin	0.036	n.d	0.13	1.16	1.75	n.d	0.24	9.80	4.87
Endrin	0.030	n.d	0.36	10.70	3.49	n.d	0.30	6.65	2.94
Endrin aldehyde	0.030	n.d	0.19	1.30	1.26	n.d	0.36	4.79	2.42
Endrin Ketone	0.032	n.d	3.78	82.17	2.95	n.d	22.24	1199.97	6.47
α -Endosulfan	0.017	n.d	9.22	710.34	8.72	n.d	0.12	3.02	3.25
β -Endosulfan	0.017	n.d	2.57	176.79	7.81	n.d	0.36	10.36	4.18
endosulfan-sulfate	0.064	n.d	1.46	17.43	1.92	n.d	2.61	79.74	3.84
HCB	0.009	0.011	0.17	2.39	2.25	n.d	0.48	13.37	3.53
methoxychlor	0.025	n.d	3.64	53.23	2.23	n.d	10.96	521.79	5.80
OCPs	-	0.0011	29.91	1043.98	0.27	n.d	60.16	1914.10	4.02

Endosulfan was the most dominant group accounting for 44.42% of the total OCPs, followed by DDTs with 17.60%, Drins (15.75%), methoxychlor (12.17%), HCHs (6.08%), Chlordane related-compounds (3.53%) and HCB (0.55%) in urban areas (Fig. 2). In agricultural areas abundances were in the order: Drins (39.46%) > DDTs (29.94%) > methoxychlor (18.22%) > Endosulfan (5.12%) > HCHs (5.06%) > Chlordanes (1.40) > HCB (0.79%).

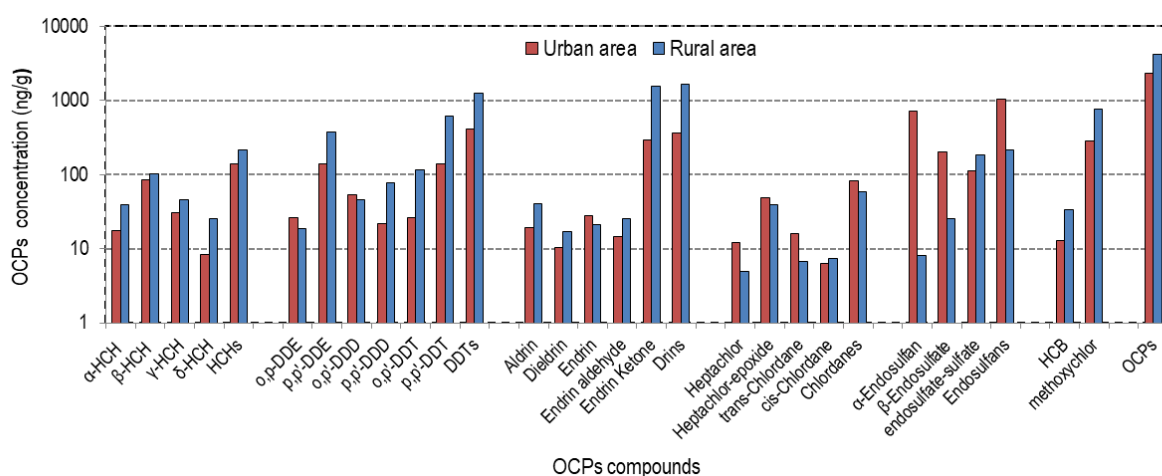


Fig. 2. Variation of individual and total (HCHs, DDTs, Drins, Chlordanes, Endosulfans, OCPs) OCP concentrations in urban areas and agricultural soils. Note the logarithmic scale applied on the Y axis.

3.1.2. Total DDT and derived metabolites

The total concentration of DDTs ranged from nd to 56.97 ng/g (mean = 5.26 ng/g - urban), and from nd to 632.95 ng/g (mean = 18.01 ng/g - rural). The highest DDTs concentrations in urban area, ranging from 24.82 to 56.97 ng/g, were found in the Sarno Basin (Campania), Apulia (Bari and Foggia) and Abruzzo (Fig. 3A). In contrast, the highest DDTs concentrations of rural areas, ranging from 400 to 628 ng/g, were found around Naples (Campania) where the vast majority of intensive agricultural land is located (Fig. 3B). In particular, total DDTs concentration presented a significantly skewed distributions as well as clear “outliers” (Figs. 3C and 3D). The latter, observed in rural areas around Naples (Campania – Fig 3D), can be considered as anomaly concentrations, which could be linked to the input of DDT through agricultural activities. Campania and Apulia are well known for their large vineyards and olives plantations on their hills and along coastal areas (Costantini and Dazzi, 2013; ISPRA, 2014a), and high DDTs residues may originate from agricultural activities in these areas. As a general observation, urban areas for this study showed lower DDTs residues compared to those reported in similar studies such as that on Beijing urban park soils (Li et al., 2008). On the other hand, some rural

areas revealed much higher DDTs residues compared to those reported in counterparts studies (Table 3).

Table 3. Total OCP concentrations (ng/g dry weight) in topsoil of the survey area compared to those found in other studies in the recent literature.

Locations	Characteristic	DDTs	HCHs	Drins	Endosulfans	Chlordanes	References
Southern Italy	Urban soils	nd – 56.97	nd – 25.08	nd – 82.58	nd – 904.2	nd – 12.47	This study
Southern Italy	Rural soils	nd – 632.95	nd – 47.27	nd – 1214.41	nd – 93.13	nd – 14.69	This study
Northern France	Natural areas	nd – 28.6	nd – 5.06	nd – 2.26	nd – 1.84	–	Villanneau et al. 2011
Central Germany	Agriculture fields	23.7–173	4.6–11.5	–	–	–	Manz et al. 2001
Southern of Poland	Urban and rural soils	23 – 260	1.1 – 11	–	–	–	Falandysz et al. 2001
Southern of USA	Farm lands	0.10 – 1490	0.1 – 0.71	–	–	0.05 – 5.1	Bidleman et al. 2004
Zhangzhou China	Agriculture soils	0.64 – 78.07	0.72 – 30.16	–	–	–	Yang et al. 2012
Beijing (China)	Urban soils park	0.942 – 1039	0.25 – 197.0	–	–	–	Li et al. 2008
Nagaon distrcit (India)	Agriculture soils	166 – 2288	98 – 1945	–	–	–	Mishra et al. 2012

Technical DDT is made up of six congeners compounds such as p,p'-DDT, o,p'-DDT, p,p'-DDE, o,p'-DDE, p,p'-DDD and o,p'-DDD. Moreover, it contains 65-80% of p,p'-DDT, 15-21% of o,p'-DDT, up to 4% of p,p'-DDD and impurities (Metcaft, 1995). In nature, p,p'-DDE and p,p'-DDD are the two main products of dechlorination of p,p'-DDT by microorganisms and/or physico-chemical properties of soil (Pfaender and Alexander, 1972; Mackay et al., 1992). More recently, dicofol has been introduced, which is structurally similar to DDT and contains high impurity of DDT-related compounds (25% of o,p'-DDT) (Qiu et al. 2005). The ratios between the parent compound and its metabolite can provide useful information on the DDT sources. For example, a survey on the formulated dicofol in China found that the ratio of o,p'-DDT/p,p'-DDT in air (Qiu et al., 2005) and soil (Yang et al., 2008) was as high as 7.

In this study, of the various compounds, the p,p'-DDT isomer was predominant, with 34.44% (urban) and 49.43% (rural). Its ranges went from nd to 16.98 ng/g (urban) and from nd to 418 ng/g (rural) (Table 2). The p, p'-DDE isomer had the second highest percentage (34.38%) and ranged from nd to 38.58 ng/g (mean = 1.81 ng/g). This was followed by o,p'-DDD (13.18%), o,p'-DDT (6.39%), o,p'-DDE (6.36%), and p,p'-DDD (5.24%) in urban areas. On the other hand, agricultural areas presented a higher dominance of p,p'-DDT (49.43%) followed by p,p'-DDE (29.96%) > o,p'-DDT (9.32%) > p,p'-DDD (6.17%) > (o,p'-DDD (3.64%) > o,p'-DDE (1.49%).

When using the o,p'-DDT/p,p'-DDT ratio (Fig. 4A), this survey highlighted a broad range of values, from 0.0002 to 214 (mean = 3.46 – urban), and from 0.008 to 16.06 (mean = 0.74 – rural). In general, the vast majority (92.51%) of the urban and rural sampling sites displayed a o,p'-DDT/p,p'-DDT ratio below 7. However, high o,p'-DDT/p,p'-DDT ratio (above 7.0) were found in some locations, mainly within

urban areas. Therefore, results point towards a predominance of historical application of technical DDT with the exception of some potential recent use of dicofol for the above highlighted urban areas.

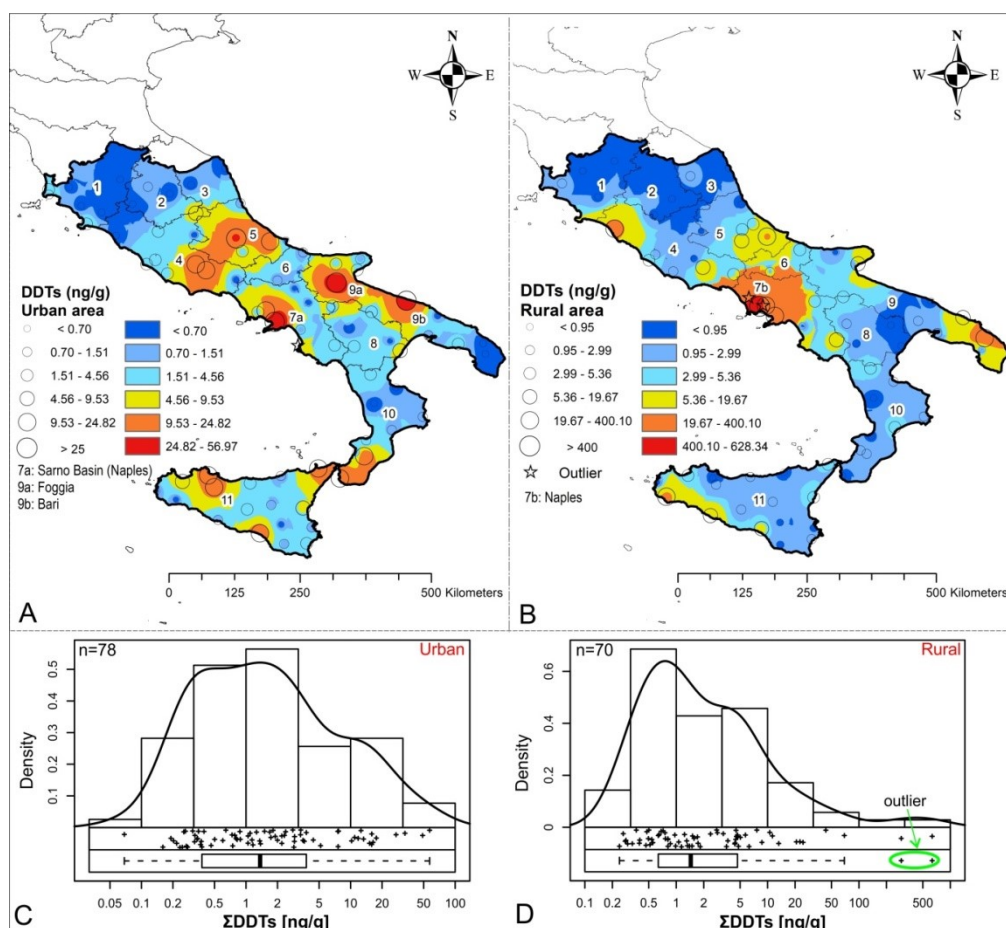


Fig. 3. Distribution of total DDTs in urban (A) and rural (B) areas. The concentration–area (C–A) fractal method was applied to set DDT ranges. Edaplots (combination of histogram, density trace, one-dimensional scattergram and Box plot) of DDTs raw data in urban (C) and rural areas (D) are displayed

Using the assumption that all p,p' -DDE and p,p' -DDD are degraded products of p,p' -DDT metabolite, the ratio of p,p' -DDT/(p,p' -DDE + p,p' -DDD) can be used to discern between historic applications of technical DDT (ratio < 1), compared to fresh or more recent applications (with ratio > 1) (Jiang et al., 2009). Results for this diagnostic ratio are again showing a significant range (Fig. 4B), from 0.0014 to 55.02 (mean = 4.02 - urban), and from 0.006 to 40.42 (mean = 2.55 – rural). In this case, less than half of the sites (47.2%) presented a ratio below 1. When using a value of 10 as arbitrary threshold for this ratio, a large number of urban areas resulted above it. It can be derived that residues of DDT for this study can be linked to a mixed contribution from historical and recent (illegal) application. The latter

was mostly highlighted in urban areas, similarly to the findings of Estellano et al. (2012), which emphasised the possible use of illegal technical DDT or dicofol in urban areas of the Tuscany region.

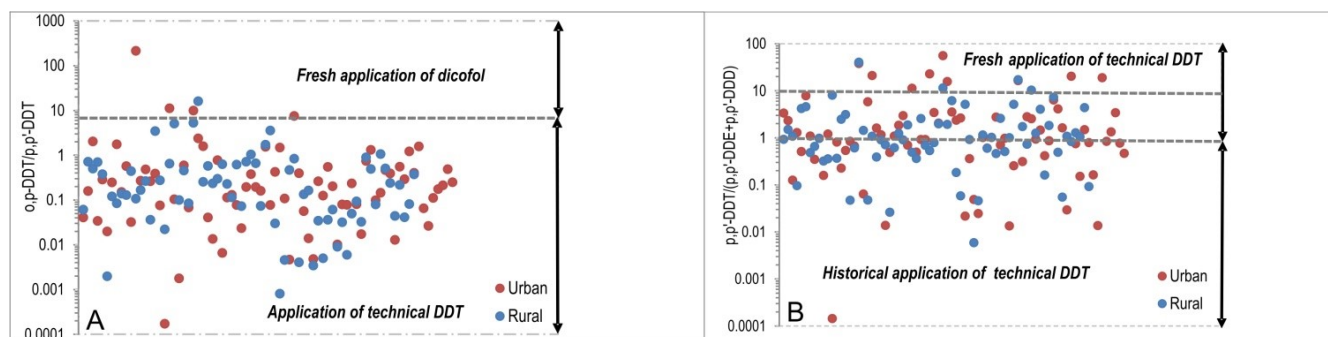


Fig. 4. Scatter diagrams of o,p'-DDT/p,p'-DDT (a) and p,p'-DDT/(p,p'-DDE + p,p'-DDD) ratios (b). The different symbology reflects whether the sites were urban (red dots) and rural nature (blue dots)

3.1.3. Total HCHs and metabolites

Total HCHs concentrations (sum of α -HCH, β -HCH, γ -HCH and δ -HCH) ranged from nd to 25.08 ng/g (mean = 1.82 ng/g – urban), and from nd to 47.27 ng/g (mean = 3.04 ng/g – rural) (Table 2). The highest values of HCHs (18.67 to 25.07 ng/g) were found in the urban areas of Bari (Apulia) (with γ -HCH isomer = 14.18 ng/g), and in the agricultural areas in the Frosinone (Latium) and Lecce (Apulia), (from 23.69 to 47.11 ng/g, with β -HCH the predominant metabolite - 20.37 ng/g) (Figs. 5A, B). Low HCHs concentrations (from nd to 2.49) were found in several areas in Tuscany, Umbria, and Marches as well as in Calabria and Sicily, whilst higher values (from 2.49 to 25.07) were found in Latium, Campania and Apulia. These HCHs spatial variations were well captured by the bimodal distributions (Figs. 4C, D) indicating the existence of two different inputs or processes controlling the patterns of HCH in the study area. No outliers were recorded, but significant departure from the mean were instead highlighted (Figs 5C, D).

The β -HCH accounted for 60.25% and 48.31% of the total HCHs, ranging from nd to 5.49 ng/g (urban), and from nd to 20.37ng/g (rural). These values are followed by γ -HCH (21.60%) > α -HCH (12.24%) > δ -HCH (5.91%) in urban, and γ -HCH (21.29%) > α -HCH (18.62%) > δ -HCH (11.78%) in rural soils. The dominance of β -HCH among HCHs isomers may be related to its resistance to degradation, and its persistence for several years in soils (Mackay et al., 1992; Calvelo Pereira et al., 2006). High residue of β -HCH isomer in rural soils of the Frosinone district (Latium) could be linked to the high contamination level of β -HCH found in the sediments of the Sacco River valley (Latium), polluted by a nearby industrial landfill percolations containing by-products of Lindane (Bianconi et al., 2010; Battisti et al., 2013).

When compared to HCHs concentrations in European soils such as those found in natural areas from northern France (Villanneau et al., 2011), in agricultural soils from central Germany (Manz et al., 2001) and rural soils from southern Poland (Falandysz et al., 2001), the findings of this study reveal higher levels in comparison. On the other hand, this study presents lower levels compared to other studies, such as those related to agricultural soils of the Nagaon District (Mishra et al., 2012) and urban park of Beijing (Li et al., 2008), which highlighted HCHs concentrations ranging from 98 to 1945 ng/g, and 0.25 to 197 ng/g, respectively.

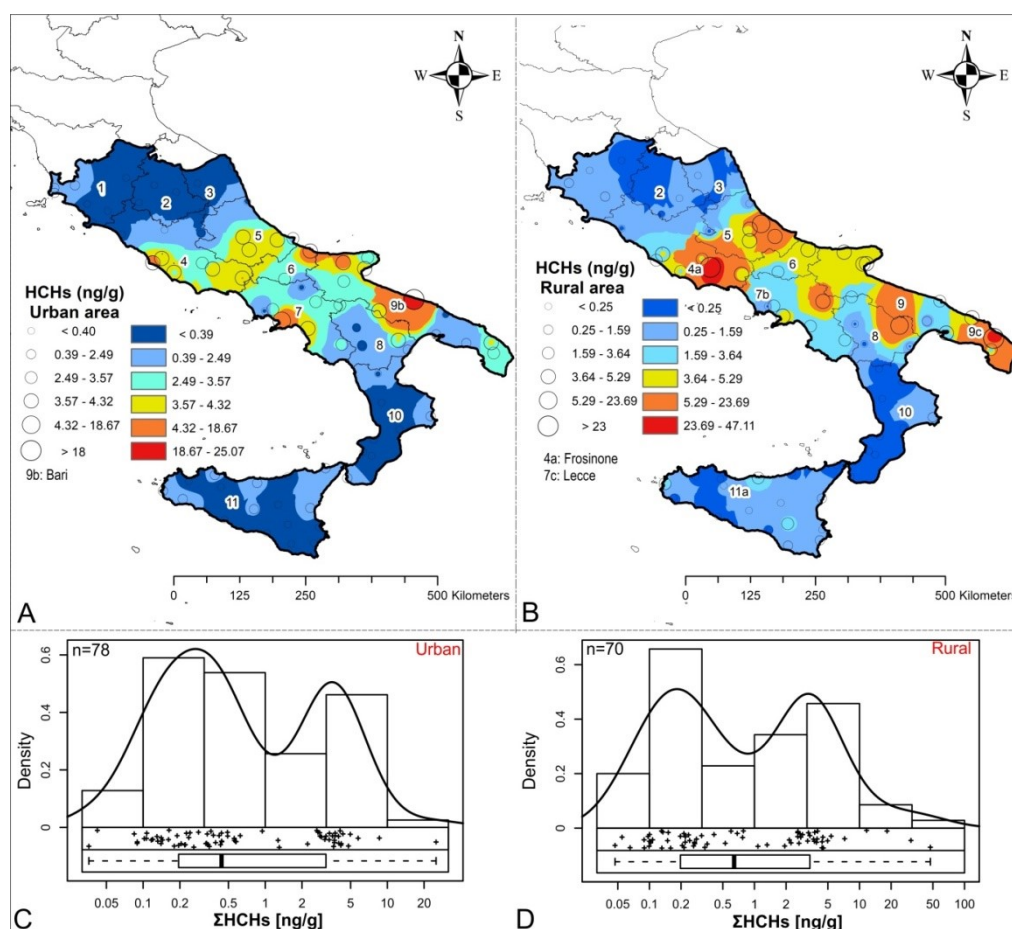


Fig. 5. Spatial distribution of HCHs in urban (A) and rural (B) areas. C–A fractal method was applied to set HCH ranges. Edaplots allow distinguishing occurrence of B “outliers” observations in the survey (C, D) areas.

Technical HCHs (60-70% α -HCH, 5-12% β -HCH, 10-12% γ -HCH, 6-10% δ -HCH and impurities) and Lindane (99% γ -HCH) are two commercial pesticides compounds that are restricted for application in Italy through European Directive in 2000 (Persistent organic pollutants amending Directive 79/117/EEC). HCH isomers have different fate and behaviour in environment. In particular, α - and γ -HCH isomers can be transformed by sunlight and through biodegradation into β -HCH, which is easily absorbed and more difficult to be evaporated from soil (Mackay et al., 1992; Calvelo Pereira et al., 2006). Studies revealed that the spatial arrangement of chlorine atoms in the β -HCH molecule protects

the compound from a microbial degradation (e.g., Walker, 1999). To distinguish application of technical HCH from a use of Lindane, the diagnostic ratio of α/γ -HCH has been successfully used (Zhang et al., 2004), with ratios from 4.64 to 5.83 being related to application of technical HCH and nearly zero for Lindane applications (Zhang et al., 2004). Results for this study highlighted α/γ -HCH ratios ranging from 0.06 to 568 (mean = 12.96 – urban), and from 0.09 to 78.19 (mean = 4.19 – rural) (Fig. 6A). A proportion of 35.2% of the samples sites presented α/γ -HCH ratio below 1, 32.9% between 1 to 4.64, 12.2 % between 4.64 and 5.83, and 9.2% a ratio above 5.83, mostly in urban areas. The 22% of the sampling sites showing a ratio of α/γ -HCH above 4.64 can possibly be linked to applications of technical DDT.

The ratios of α/β -HCH ranged from 0.002 to 822 (mean = 19.3 – urban), and from 0.005 to 180 (mean = 8.21 – rural) (Fig. 6B). Here a proportion of 52.6% of the sampling sites presented α/β -HCH ratio below 1.0. The findings seem to indicate both historical application and (illegal) recent use of technical HCH in soils of the survey area. Assessment of OCPs in air samples from the Tuscany region (Estellano et al., 2012) revealed possible illegal use of technical HCH or Lindane in some urban areas.

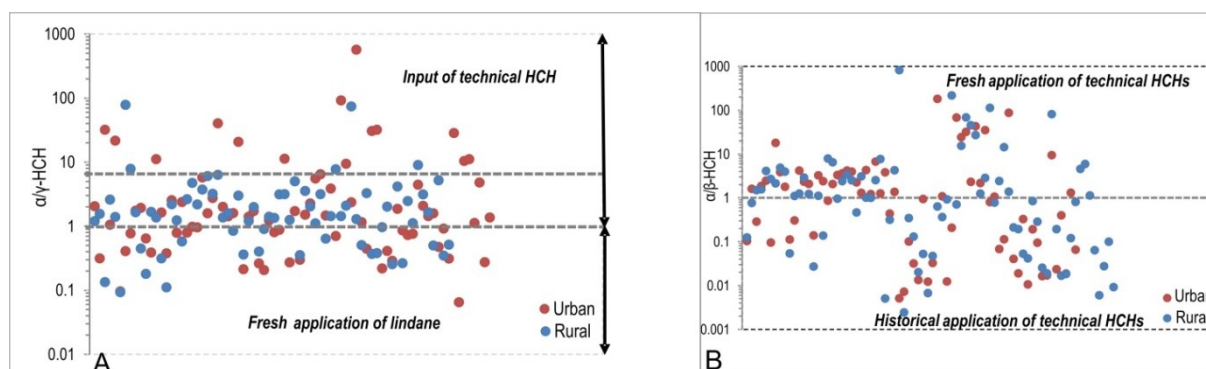


Fig. 6. Scatter diagrams of α/γ -HCH (a) and α/β -HCH (b) ratios.

3.1.4. Drins

Dieldrin, Aldrin and Endrins are collectively called Drins or Drin pesticides and were synthesized from pentadiens obtained as secondary products of petro-chemistry through the Diels-Alder reaction (Oppolzer, 1991). They were primarily used as an insecticide, as well as a rodenticide and piscicide. Total Drins (sum of Dieldrin, Aldrin, Endrin, Endrin aldehyde, and Endrin Ketone) for this study ranged from nd to 82.5 ng/g (urban) and from nd to 1212 ng/g (rural). The highest urban concentrations, ranging from 31.85 to 82.5 ng/g, were found in Apulia (Bari and Foggia) and Abruzzo, whereas rural areas in the Sarno Basin (Campania) and Lecce (Apulia) presented high Drins values, ranging from 120.2 to 1212 ng/g (Figs. 7A, B). Statistically abnormal distributions and outliers were observed both in urban areas and in rural areas (Figs 7C and D).

Among Drins, Endrin Ketone was the predominant compound accounting for 80.28% (urban) and 93.71% (rural), ranging from nd to 82.16 ng/g (urban), and from nd to 1199 ng/g (rural) (Table 2). Endrin Ketone is the final photodegradation product of Endrin and Endrin Aldehyde, and is difficult to further degrade (Fan and Alexeeff, 1999). These results may indicate that the Drins residues in soils are mainly the result of historical application across the study area. In comparison with other studies, for example with reported values from northern France (Villanneau et al., 2001), the present survey showed higher concentrations of Drins in urban and rural areas.

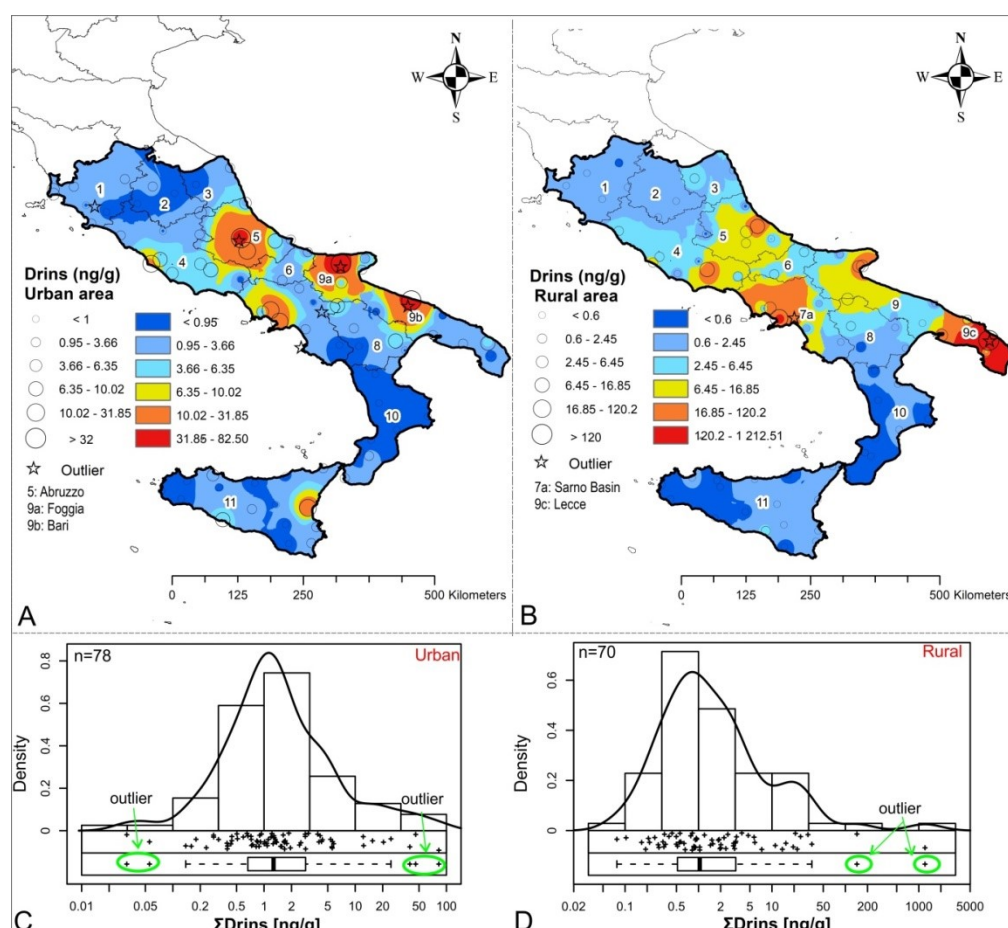


Fig. 7. Spatial distribution of Drins in urban (A) and rural (B) areas; the concentration–area (C–A) fractal method was applied to set concentration ranges. Edaplots (C, D) allow distinguishing occurrence of outlier observations in urban and rural areas.

3.1.5. Chlordanes related-compounds

Technical Chlordane is generally used for insecticides, herbicide and termiticides, and is a mixture of more than 140 related compounds (Dearth and Hites, 1991). Sixty to 85% of technical chlordane is made up by stereoisomers cis- and trans-chlordane with a mixture of minor compounds such as Heptachlor, Heptachlor epoxide, cis and trans-nonachlor (Parlar et al., 1979). In this study, total

concentrations of Chlordane related compounds (sum of cis-chlordane, trans-chlordane, Heptachlor and Heptachlor-epoxide) ranged from nd to 12.46 ng/g (mean = 1.05 ng/g – urban), and from nd to 14.68 ng/g (mean = 0.84 ng/g – rural). High urban concentrations of Chlordanes were found in Campania and Bari (Apulia), Palermo (Sicily), Grosseto (Tuscany), ranging from 10.03 to 12.46 ng/g, whilst large rural Chlordanes values were found in Tuscany, Campania (Naples) and Sicily showed, ranging from 6.11 to 14.68 ng/g (Figs. 8A, B). These results were confirmed by the presence of statistically abnormal distributions of Chlordane related compounds and, by one outlier (anomaly - Figs. 8C, D).

Among Chlordane related compounds, Heptachlor epoxide was the prevalent with 58.37% (urban) and 67.56% (rural). Heptachlor epoxide is explained as an oxidation and biodegradation product of Heptachlor which has been used in the past for killing insects in households, buildings, and on food crops, especially corn (Pornomo et al., 2013). Chlordane related compounds have been banned in 1988 (ATSDR, 1995). Thus, large Heptachlor epoxide concentration, and mean values of Heptachlor/Heptachlor epoxide ratio equal to 0.23 (urban) and 0.14 (rural) point towards historical application of the commercial Chlordane. However, when compared to similar studies, such as that conducted by Bidleman et al. (2004) in farmland of the Southern of USA (Chlordane related compounds concentration ranging from 0.05 to 5.1 ng/g), the results from this study seem to suggest extremely extensive applications made in some parts of the studied area.

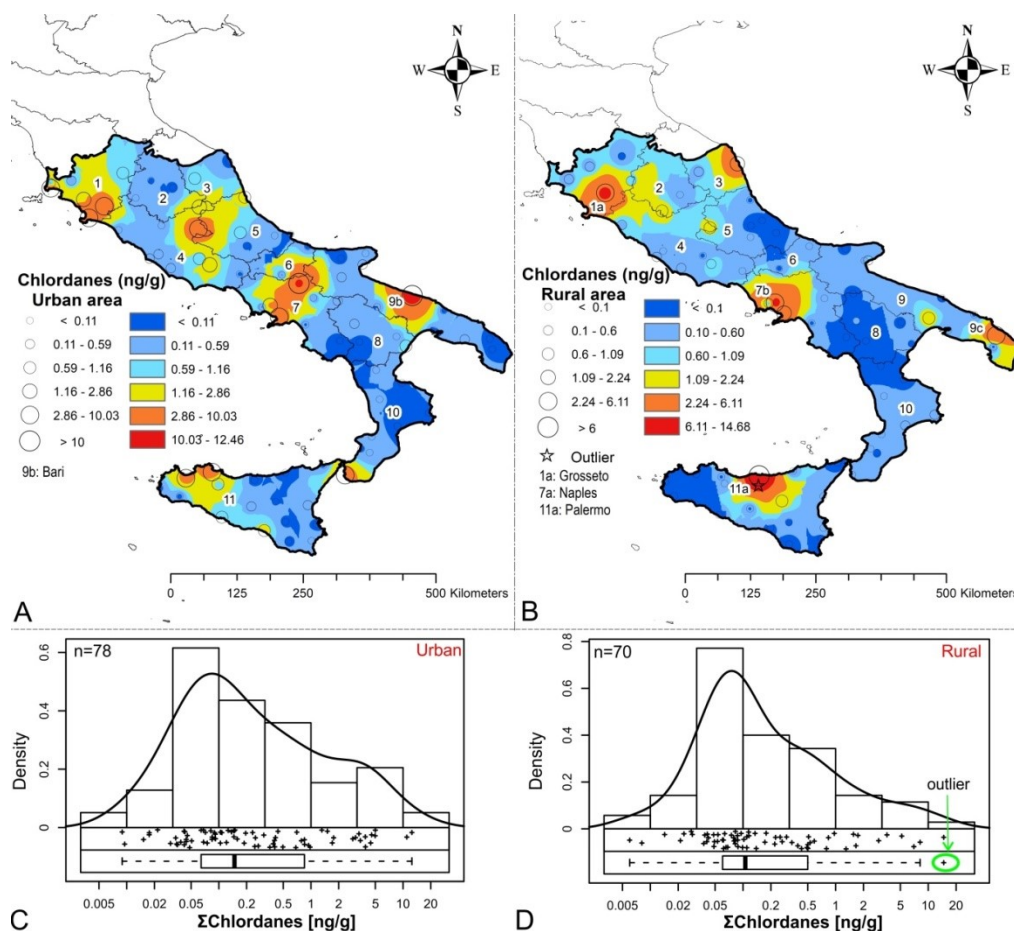


Fig. 8. Distribution of total Chlordane-related compounds in urban (A) and rural (B) areas. Edaplots (C,D) reveal abnormal distribution of the dataset and occurrence of an outlier rural area.

3.1.6. Endosulfans

Endosulfan is a cyclodiene pesticide used worldwide to control pests in non-food crops (cotton, tobacco, timber, and ornamental plants), food crops such as vegetables, fruits, corn, and cereals and a control a wide variety of insects and mites (ATSDR, 2000). Italy is the second consumer of Endosulfan in European Union with 20% of the total volume, after Spain (Endosulfan Preliminary Dossier, 2003). Technical endosulfan was globally banned under the Stockholm Convention (2011) because of its threats to human health and the environment. Endosulfan is made up α - and β -endosulfan isomers that are fairly resistant to degradation and persistent in the environment. Endosulfan sulfate is the degradation product of Endosulfan, and it is a more hydro-soluble metabolite and susceptible to photolysis (Cerrillo et al., 2005).

In this survey, total Endosulfan (sum of α -endosulfan, β -endosulfan, and endosulfan sulfate) ranged from nd to 904.21 ng/g (mean = 13.25 ng/g) accounting for 44.32% of the total OCPs in urban areas, and from nd to 92.99 ng/g (mean = 3.08 ng/g) accounting for 5.12% of total OCPs in rural area. High Endosulfan concentrations were found in the urban area of Bari (Apulia), ranging from 71 to 904.21ng/g

and in rural areas of Lecce (Apulia) from 55.32 to 92.99 ng/g (Figs. 9A, B). These values are extremely large if compared to those found in natural areas of the Northern France (ranging from nd to 1.84 ng/g - Villanneau et al., 2001). Statistical distributions showed both outliers as well as abnormal behaviour of Endosulfans concentrations (Fig. 9C, D), which could be associated with the diverse chemical processes that may affect endosulfan compound behaviour in soils medium. Since α -endosulfan decomposes more easily than β -endosulfan in soil, the ratio of α/β -endosulfan < 2.33 may be used to judge the age of their residues in soil (Jennings and Li., 2014; Jia et al., 2010). In urban areas, α -endosulfan isomer constituted 69.59% of the total endosulfan followed by β -endosulfan with 19.36% and endosulfan sulphate (11.05%), and the ratio of α/β -endosulfan ranged from 0.05 to 312.9 (mean = 22.44). Endosulfan sulphate was the predominant compound in rural areas (84.58%), followed by β -endosulfan and α -endosulfan, and the ratio of α/β -endosulfan ranged from nd to 40 (mean = 1.59). These results strongly suggest a recent (illegal) use of technical endosulfan in urban areas, especially in Apulia. In contrast, results for rural areas seem to point to historical application. The relatively recent restriction of technical endosulfan (Stockholm Convention, 2011) and its uses in Italy until December 2007 may explain why it was still found in high proportion in the soils of the survey area (Pozo et al., 2016; Qu et al., 2017).

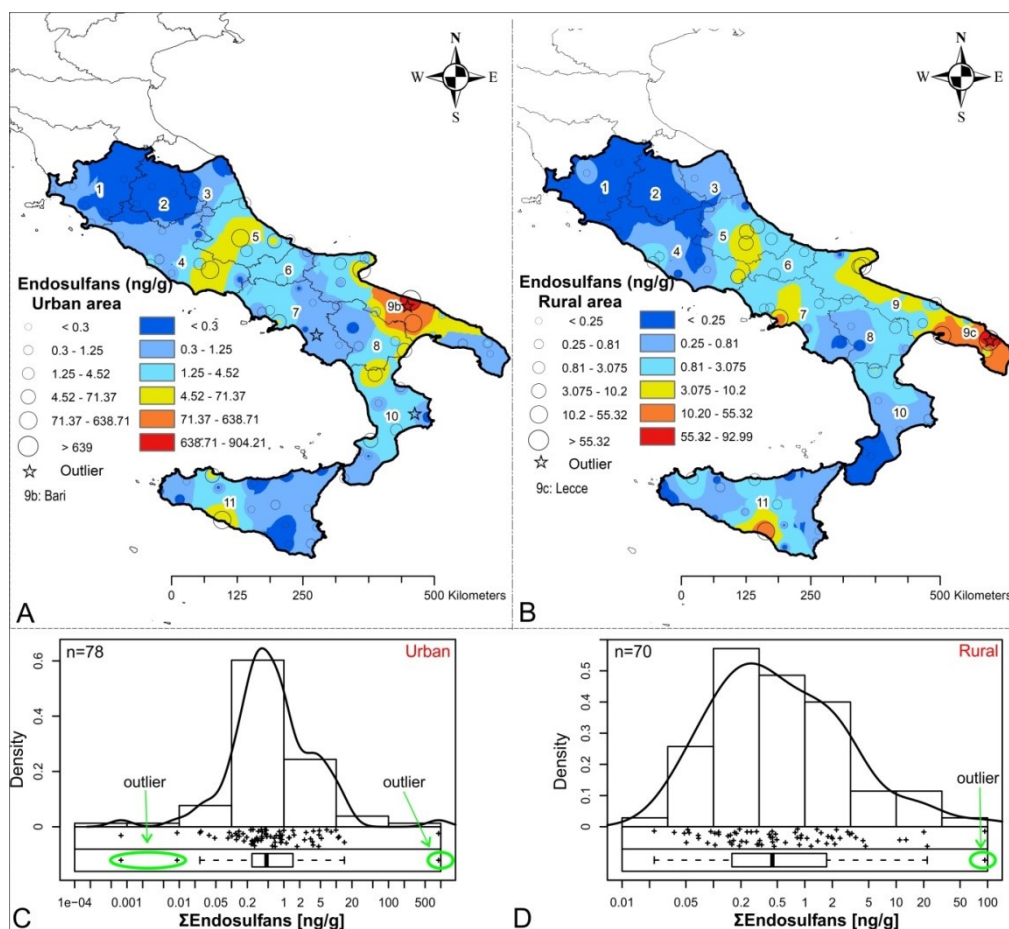


Fig. 9. Distribution of concentration of Endosulfan in urban (A) and rural (B) areas. Edaplots (C,D) reveal abnormal distribution of the dataset and occurrence of low and high outliers in urban and rural areas.

3.1.7. HCB and methoxychlor

HCB was listed among the first group of persistent OCPs compounds in the Stockholm Convention (Stockholm Convention, 2005), even though it has been restricted since 1985 in the European Union countries (Barber et al., 2005). It has been used as fungicide to control bunt on wheat, and seed treatment of onions and sorghum (Courtney, 1979). The values of HCB in this survey ranged from 0.01 to 2.39 ng/g (mean = 0.16 ng/g – urban), and from nd to 13.37 ng/g (mean = 0.47 ng/g – rural). HCB made up 0.55% (urban) and 0.79% (rural) of the total OCPs concentrations. Several studies reported that HCB is still used as a by-product or impurity in several chemical compounds, including chlorinated pesticides such as Lindane (Pacyna et al., 2003; Barber et al., 2005). Pearson correlation coefficient between HCB and γ -HCH compounds showed a slight correlation ($r=0.44$), which may suggest that HCB could be partially related to input of technical HCH or Lindane in the study area.

Most methoxychlor enters the environment when it is applied to forests, agricultural crops, and farm animals as insecticide (US EPA, 1991). It is one of the few organochlorine pesticides that has undergone an increase in its use since the ban on DDT, but methoxychlor was finally listed as banned

OCPs pesticides by the United Nations Environmental Program (UNEP) (Stockholm Convention, 2011). In this study, the concentrations of methoxychlor ranged from nd to 53 ng/g (mean = 3.64 ng/g – urban) and from nd to 521 ng/g (mean of 10.96 ng/g – rural). When compared to other studies, the mean concentration of the methoxychlor (10.96) found in rural areas is comparable to that from agricultural soils of central China (Zhou et al., 2013), but bigger than that found in southern Mexico (Cantu-Soto et al., 2011) and soils from the hilly areas of Nepal (Yadav et al., 2017).

3.2. Compositional Biplot and robust Factor analysis

Compositional biplots explained 66.9% (PC1-PC2) and 61.5% (PC1-PC3) of the variability (Fig. 10).

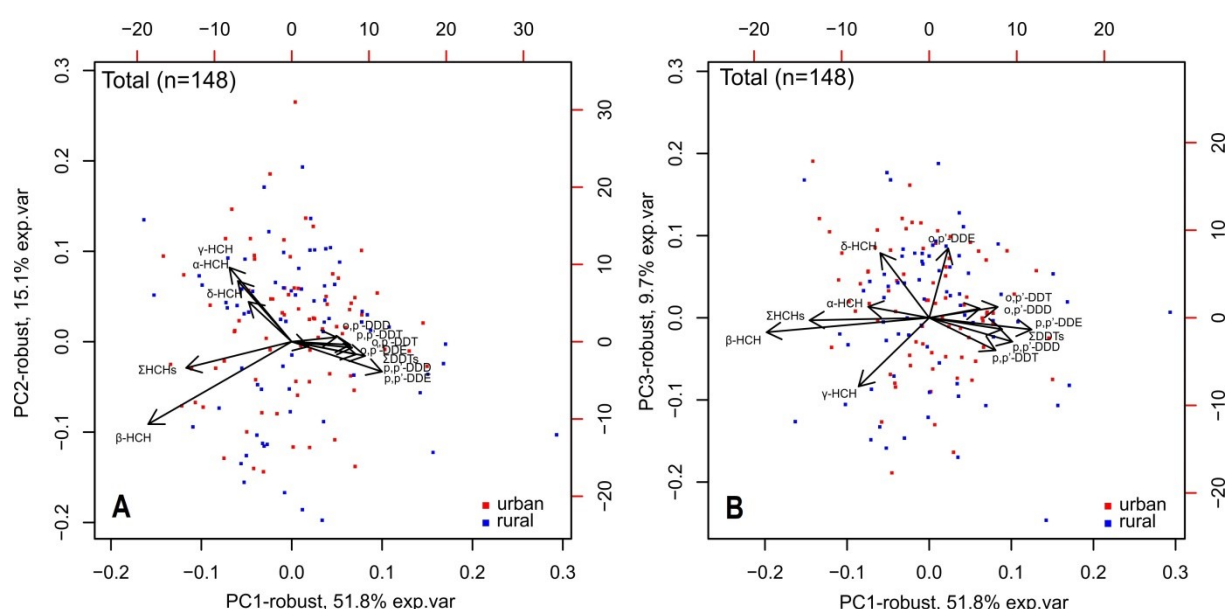


Fig. 10. Robust biplots for the first and second principal components (A) and for the first and third principal components (B) based on DDTs and HCHs investigated compounds

The sum of DDTs (p,p'-DDT, o,p'-DDT, p,p'-DDE, o,p'-DDE, p,p'-DDD, and o,p'-DDD) may be considered a variables association (A), due to the vicinity of their vertices and their rays pointing to the same direction. Among the associations, DDTs and p,p'-DDE displayed the highest vector length (communality - Fig. 10A). It was possible to discriminate DDTs variables and highlights two sub-groups based to their chemical structures (Fig. 10B):

- A₁, formed by o,p'-DDT and o,p'-DDD (superimposition of vertices and low communalities, and
- A₂, formed by p,p'-DDT, p,p'-DDE, p,p'-DDD (proximity and high communality).

Furthermore, the o,p'-DDE variable presented a high length vector and it was separated from other DDTs metabolites. These findings point towards a significant interrelationship between DDT isomers.

High lengths of p,p'-DDTs metabolites could be associated with a dominant input of technical DDT in soils of the study area (Qu et al. 2016). Low communalities and superimposition of the o,p'-DDT and o,p'-DDD variables may also be associated with the use of dicofol which contains more o,p'-DDTs metabolites (Qi et al., 2005). Moreover, the high length of the o,p'-DDE vector and its disassociation to others DDTs isomers may illustrate its specific behaviour in soil based on its specific physico-chemical properties (Solubility, partition coefficient, and vapour pressure) (Pfaender and Alexander, 1972).

Another variables association (B) made by α -HCH, γ -HCH and δ -HCH was highlighted (Fig. 10A). This was also characterized by the vicinity of their vertices with their vectors being superimposed to one another. This configuration can be associated to a similar behaviour of these three isomers in the soils of the study area. In fact, α -HCH, γ -HCH and δ -HCH metabolites are the main compounds related to the commercial technical HCH (Senthil Kumar et al., 2001), which is reflected by how α -HCH and δ -HCH vectors are geometrically symmetric respect to the HCHs vector (Fig. 10B). This spatial configuration illustrates their similar behaviour in the soils of the study area. The β -HCH variable is marked by a high communality, and its vector forms 90° with the (B) variables association (Fig. 10A). These results seem to suggest that the β -HCH compound has a different fate and behaviour in soils compared to other HCH metabolites. Moreover, the high communality and disassociation of β -HCH respect to other HCH metabolites could explain its persistence to degradation and its accumulation in soil of the study area (Jiang et al., 2009).

Factor analysis was performed to determine the correlation between DDT and HCH isomers which further revealed the possible sources of these compounds. Factor loadings and total variances of individual OCPs were computed to facilitate the interpretation (Table 4). DDT and HCH metabolites of the three-factor model were separated by positive and negative loadings and sorted in descending order:

- Urban sites

F₁: o,p'-DDT, p,p'-DDD, - (γ -HCH, α -HCH)

F₂: p,p'-DDD, - (δ -HCH, β -HCH)

F₃: p,p'-DDT, - (o,p'-DDE)

- Rural sites

F₁: p,p'-DDD, o,p'-DDT, o,p'-DDD, - (γ -HCH, α -HCH)

F₂: p,p'-DDE, o,p'-DDE, - (β -HCH)

F₃: o,p'-DDD, - (p,p'-DDT)

Table 4. Varimax-rotated factor (three-factor model) using 78 topsoil samples from urban areas and 70 samples from agricultural soils; bold entries: loading values over |0.50|

Variables	Urban areas			Rural areas		
	Factors			Factors		
	F1	F2	F3	F1	F2	F3
α -HCH	-0.69	0.12	-0.36	-0.80	-0.08	0.19
β -HCH	0.10	-0.83	-0.12	-0.15	-0.77	-0.06
γ -HCH	-0.68	0.04	-0.19	-0.71	-0.26	0.03
δ -HCH	-0.25	-0.64	0.17	-0.24	-0.43	0.31
o,p'-DDE	-0.33	0.04	-0.75	-0.45	0.53	0.33
p,p'-DDE	0.04	0.81	0.12	0.16	0.81	-0.30
o,p'-DDD	0.45	0.33	0.47	0.64	0.15	0.62
p,p'-DDD	0.76	0.19	0.03	0.81	0.28	-0.10
o,p'-DDT	0.84	0.13	-0.13	0.79	0.03	-0.09
p,p'-DDT	-0.08	0.07	0.79	0.30	0.25	-0.81
Eigenvalues	2.60	1.96	1.64	3.77	1.675	1.23
Total variance in %	26.04	19.58	16.45	33.31	19.6	13.81
Cum. of total variance	26.04	45.63	62.09	33.31	52.92	66.81

Factor scores values for F1, ranging from -3.78 to 1.64 (urban), and -2.75 to 2.94 (rural), were plotted to represent their spatial distribution (Fig. 11). High urban factor score values (ranging from 1.26 to 1.62), associated mainly with o,p'-DDT (0.84) and p,p'-DDD (0.76) compounds, were found in Frosinone (Latium), in Foggia (Apulia), in southeastern coastal area of Calabria and in the Sicily region (Palermo and Gela) (Fig. 11A). High factor loading of the o,p'-DDT isomer (0.84), explained by its dominance in urban soils, may be attributed to the application of dicofol, containing high o,p'-DDT residue. Further increase of p,p'-DDD (0.76) isomer in these areas was relatively significant and might be related to degradation processes of DDT compounds. Low urban factor scores (< to -1.60), tied to α -HCH (-0.69) and γ -HCH (-0.68), were mainly observed in Calabria and Marches. These are potentially related to application of technical HCH. The physico-chemical properties of α -HCH and γ -HCH are similar, showing a relatively easy degradation in soils (Mackay et al., 1992; Calvelo Pereira et al., 2006). The highest rural factor score values (ranging from 1.52 to 2.94), associated with p,p'-DDD (0.81), o,p'-DDT (0.79) and o,p'-DDD (0.64) compounds, were found mainly along the coasts (Latium, Campania - Naples,- Calabria, and southern Sicily – Fig. 11B), where intensive agriculture activities such as those carried out in vineyards and olive plantations occur (Corona et al., 2012; ISPRA, 2014a). The higher loading of p,p'-DDD compound can be associated to historical applications of DDT together with a more recent application of dicofol, illustrated by occurrence of o,p'-DDT, and o,p'-DDD isomers in these

areas. This is partially in line with the results of Qu et al. (2016) which have indicated that DDT residues in the Campania plain are mainly the result of historical application.

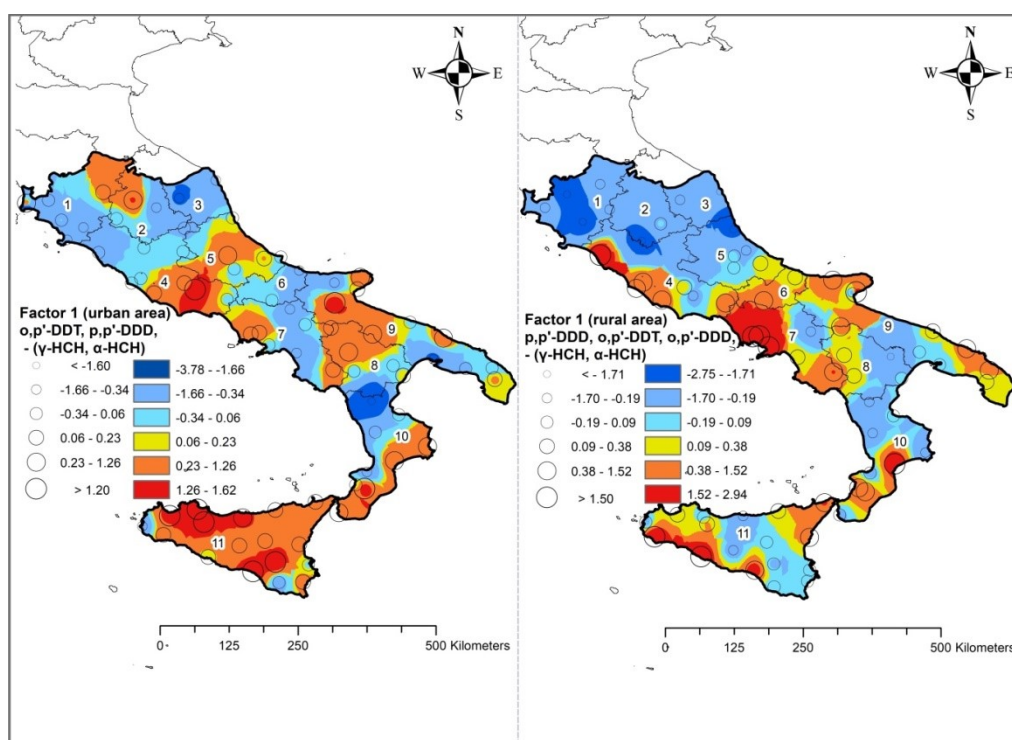


Fig. 11. Dots and interpolated factor score maps of factor 1 in urban areas (A) and rural soils (B). Factor score value ranges are based on C–A (concentration–area) fractal plot after a min–max normalisation.

F2 factor score values ranged from -2.56 to 2.26 (urban), and -2.16 to 1.92 (rural) (Fig. 12). The urban areas of Grosseto (Tuscany), northern Campania and Taranto (Apulia) displayed the highest factor scores (> 1.90) corresponding to the p,p'-DDE (0.81) isomer (Fig. 12A). These results can be associated to historical application of technical DDT because p,p'-DDE is a degradation product of p,p'-DDT isomer.

The highest F₂ factor scores (ranging from 1.77 to 2.75) were related to p,p'-DDE (0.81) and o,p'-DDE (0.53), and were found in most rural sites in the northern part of the study area and in Naples (Campania). This may be attributed to historical application of technical DDT, because high factor scores of DDE metabolites are matching their degradations and fate in situ. Low factor scores values (ranging from -2.16 to -1.39) corresponded to β -HCH (-0.77) and were found in rural sites of Frosinone (Lazio) (Fig. 12B). This might be related to the dominance or specific behaviour of the β -HCH metabolite in soils of this area. As previously mentioned, high level of the β -HCH isomer was found in soils and sediments from the Sacco River valley (Frosinone), which are polluted by the release of industrial landfill percolations containing by-product of Lindane.

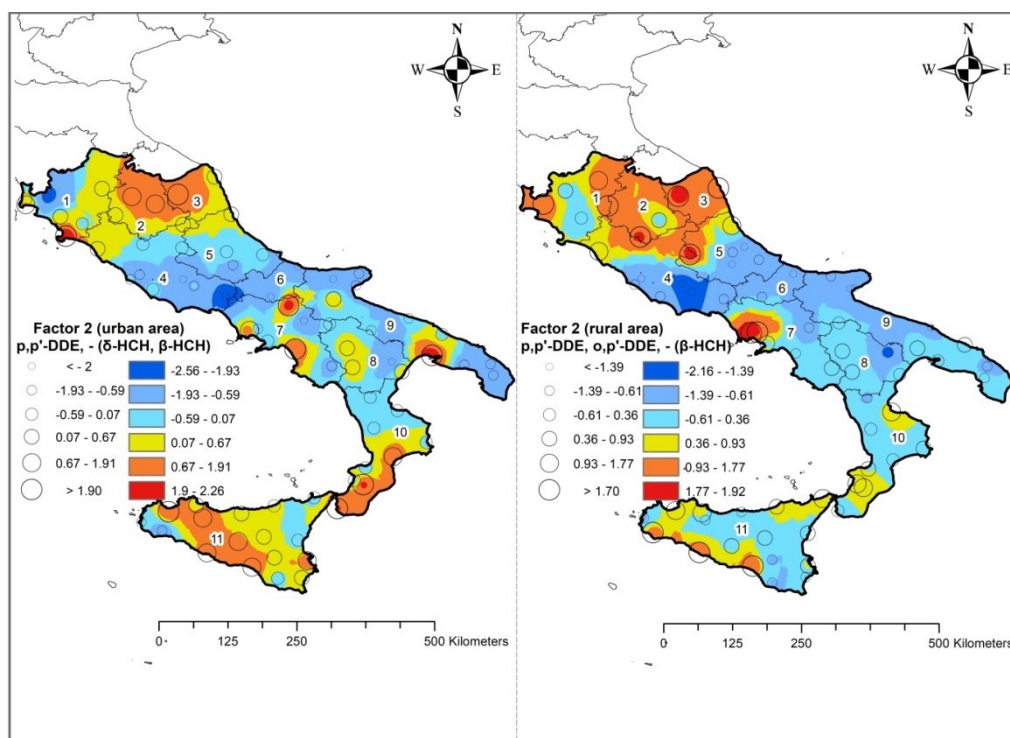


Fig. 12. Dots and interpolated factor score maps of factor 2 in urban areas (A) and rural soils (B). Factor score value ranges are based on C–A fractal plot after a min–max normalisation.

Values for the F_3 factor score, ranged from -2.83 to 2.05 (urban) and -3.93 to 2.66 (rural) (Fig. 13). The highest urban factor score values (> 1.23) corresponded to p,p'-DDT (0.79) and were found in Palermo (Sicily), Naples (Campania) and Tuscany (Fig. 13A). This is likely to be related to recent illegal application of technical DDT, which is confirmed by the dominance of the p,p'-DDT isomer in soils of these areas, similarly to what Qu et al. (2016) have found in soils of the Campania plain. Pozo et al. (2016) also highlighted recent use of technical DDT in Palermo (Sicily), whilst an assessment of OCPs pollution sources in urban air of Tuscany revealed possible illegal use of commercial technical DDT (Estellano et al., 2012). Low factor score values (ranging from -2.83 to -1.85) were found in Tuscany, and corresponded to o,p'-DDE. We might preclude that the occurrence o,p'-DDE isomer in Tuscany may related to unknown synthetic chemicals. A follow up study in this region may give reasons of the occurrence of this metabolite in Tuscany soils.

The highest rural factor score values (ranging from 2.09 to 2.66) corresponded to o,p'-DDD in Calabria (Cosenza) and Tuscany (Fig. 13). Low values (ranging from -3.93 to -1.67) were found in rural areas in Basilicata, revealing dominance of p,p'-DDT. These results point towards a mixed input of DDT residues through recent use and historical application in rural areas of Calabria.

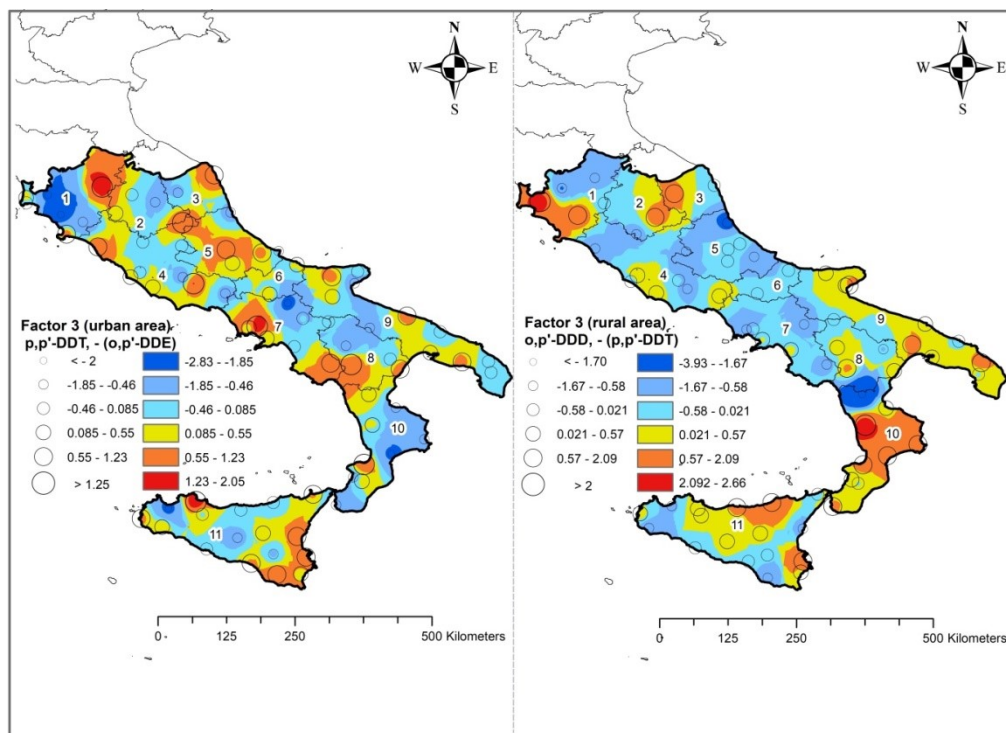


Fig. 13. Dots and interpolated factor score maps of factor 3 in urban areas (A) and rural soils (B). Factor score value ranges are based on C–A fractal plot after a min–max normalisation.

3.3. Contamination assessment

The SoQI index was used to represent the degree of contamination, and therefore concern, of the studied area (Fig. 14).

SoQI values in soils of urban and rural areas of the Tuscany, Umbria, Marche, and Molise are equal to 100 (Fig. 14A, B). This is associated to very low contamination levels, where none of the sampling sites (both in urban and rural soils) presented concentration beyond the threshold values established by Italian environmental legislation (D. Lgs. 152/2006). Similarly, low levels of concern were observed for urban and rural area of Basilicata and Calabria (Fig. 14A, B).

Urban soils in Campania (Naples, Sarno Basin), Abruzzo, and Apulia (Foggia and Bari) presented SoQI ranging from 50 to 70, corresponding to a medium contamination level. In addition, rural soils in Latium (Frosinone and Civitavecchia), Abruzzo, and Apulia (Taranto and Manfredonia) also showed the same SoQI (50 – 70).

The lowest urban SoQI value (46.2), associated to a high contamination level, was found in urban soils of the metropolitan area of Foggia. This could be related to the use of OCPs pesticides against pests in urban gardens. For rural areas, instead, the lowest SoQI values (ranging from 30 to 50) were found in soils of Campania (Naples metropolitan area and Sarno Basin) and Apulia (Lecce). This further confirms the observations made previously which linked high contamination levels with intensive agricultural activities such as those occurring in vineyards and olive plantation along the coastal areas.

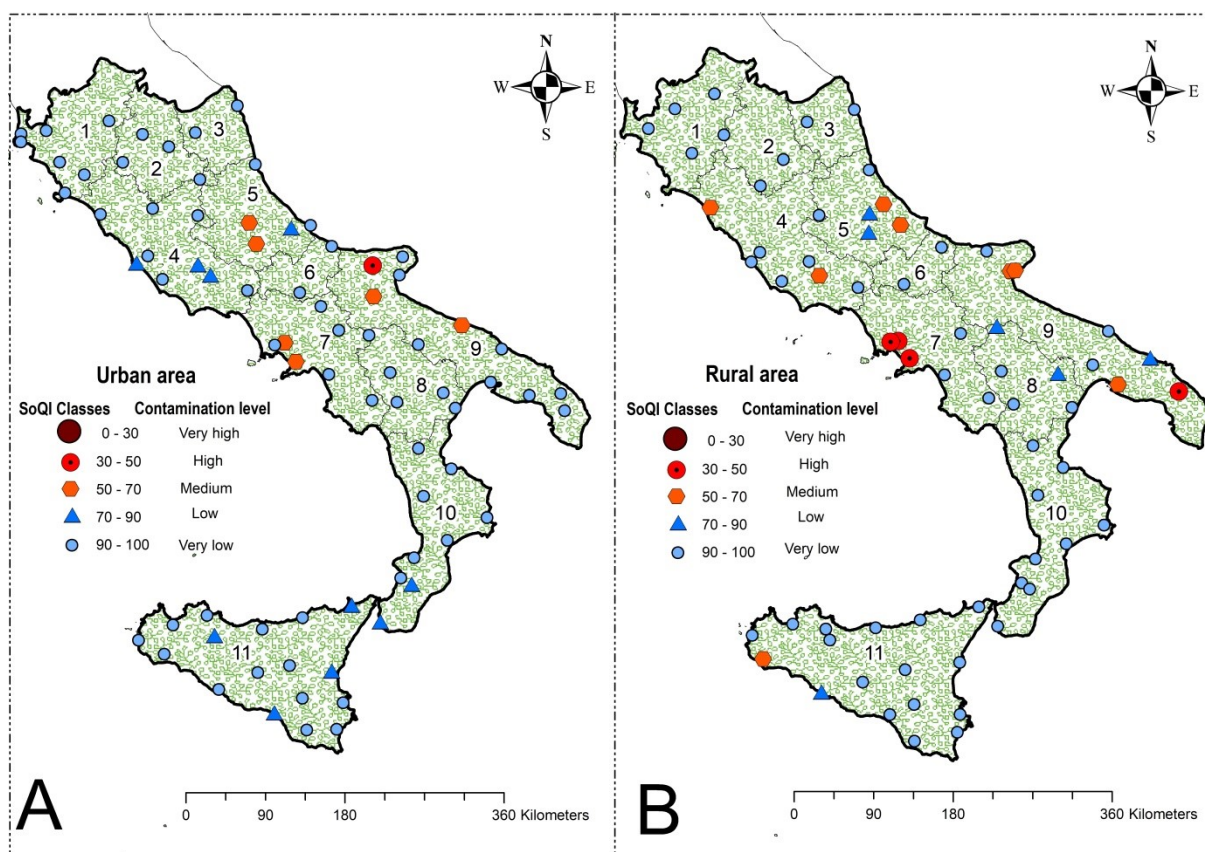


Fig. 14. Spatial distribution of SoQI contamination levels in urban (A) and rural areas (B) in the 11 regions of the study area.

4. Conclusion

This study presented the results of a regional survey of OCPs compounds conducted in urban and rural soils of 11 administrative regions from central and southern Italy, as part of an on-going project aiming to cover the entire Italian territory. The main findings revealed the concentration of 24 OCPs, ranging from nd to 1043 ng/g (mean = 29.91 ng/g – urban), and from nd to 1914 ng/g (mean of 60.16 ng/g – rural). In particular, high DDTs concentrations were mostly shown in urban and rural soils of Campania and Apulia. Enrichment of HCH was also highlighted in the central regions of the study area, with relatively lower values in the north and southern parts. Furthermore, Endosulfan related compounds and Methoxychlor were found to be 42.32% and 12.17% of the total OCPs in urban areas, respectively, which are likely to be related to recent applications, particularly in Apulia.

Diagnostic ratios of DDTs residues clearly unveiled a dominance of historical application of these compounds in soil, but also a minor (yet significant) more recent illegal use of technical DDT or dicofol mainly in urban areas. A mixed application of HCH was also highlighted, with residues both from historical and recent applications. On the other hand, the different compositions of Drins and

Chlordanes related compounds emphasized that the residues of these compounds are mainly related to a historical application. At the same time, recent applications of Endosulfan residues in urban areas were suggested, together with an historical use of this compound in agricultural soils. Unfortunately, the Italian environmental legislation has not established to date any guideline (threshold) values with regards to Endosulfan residues in soil (see D. Lgs. 152/2006), failing to recognise their potential threat to human health.

These results were also backed up by the findings of the multivariate computations performed on DDT and HCHs residues, pointing out that (1) DDT and HCHs residues could be mainly related to historical but also more recent (illegal) application; (2) occurrence of DDTs residues in soils of the Campania region could be related to historical application of technical DDT; (3) indirect evidence of illegal 'fresh' application of DDT were identified in urban areas of Tuscany, Sicily and Campania; (4) HCH levels in Latium (Frosinone) rural areas could be related to β -HCH metabolite in the anomalous sediments of the Sacco River valley, affected by nearby industrial landfill percolations.

This study should be considered as a first stepping stone (as a regional survey) towards a major investigation on the main sources and levels of OCPs throughout the Italian territory. As such, it is envisaged that the findings will contribute to build OCPs baseline and drive towards an entire coverage of the Italian territory. The survey, which is currently progressing in the remaining 9 regions of northern Italy, has highlighted areas with high concentrations of some OCPs, which can be in part explained by recent (illegal) applications. Even though it was not the scope of this study, results study highlighted some potential human health concerns, which need addressing urgently. Given the associated human health risks and the potential wider implications for the environment, these results strongly point towards follow-up studies to be held in areas of higher contamination levels (Naples and Sarno Basin as well as Foggia and Lecce regions), with a larger number and higher density of soil and air samples. It is also hoped that similar studies will build science-based evidence to be fed back at institutional level for more adequate and comprehensive regulations and, in the long-term, for a full ratification of the Stockholm Convention.

References

- Abraham, G.M.S., Parker, P.J., 2008. Assessment of heavy metal enrichment factors and the degree of contamination in marine sediments from Tamaki Estuary, Auckland, New Zealand. *Environ. Monit. Assess.* 136 (1–3), 227–238.
- Adamo, P., Iavazzo, P., Albanese, S., Agrelli, D., De Vivo, B., Lima, A., 2014. Bioavailability and soil-to-plant transfer factors as indicators of potentially toxic element contamination in agricultural soils. *Sci. Total Environ.* 500–501, 11–22.
- Adamo, P., Zampella, M., 2008. Chemical speciation to assess potentially toxic metals (PTMs) bioavailability and geochemical forms in polluted soils. In: De Vivo, B., Belkin, H.E., Lima, A. (Eds.), *Environmental Geochemistry: Site Characterization, Data Analysis and Case Histories*. Elsevier, pp. 175–212.
- Agapkina, G. I., Chikov, P. A., Shelepchikov, A. A., et al. 2007. Polycyclic aromatic hydrocarbons in soils of Moscow. *Vest. Moscow University, Series 17: pochvoved* (Vol. 3, pp. 38–46).
- Agarwal, T., Khillare, P. S., Shridhar, V. 2006. PAHs contamination in bank sediment of the Yamuna River, Delhi, India. *Environmental Monitoring and Assessment*, 123, 151–166.
- Agarwal, T., Khillare, P. S., Shridhar, V., Ray, S. 2009. Pattern, sources and toxic potential of PAHs in the agricultural soils of Delhi, India. *Journal of hazardous materials*, 163, 1033–1039.
- Agterberg, F.P., 2001. Multifractal simulation of geochemical map patterns. In: Merriam, D.F., Davis, J.C. (Eds.), *Geologic Modeling and Simulation. Computer Applications in the Earth Sciences*. Springer, Boston, MA, pp. 327–346. http://dx.doi.org/10.1007/978-1-4615-1359-9_17.
- Aichner, B., Glaser, B., Zech, W. 2007. Polycyclic aromatic hydrocarbons and polychlorinated biphenyls in urban soils from Kathmandu, Nepal. *Organic Geochemistry*, 38, 700–715.
- Aitchison, J., 1986. *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman and Hall Ltd (reprinted 2003 with additional material by The Blackburn Press), London, UK (1986). 416pp.
- Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J.A., Pawlowsky-Glahn, V., 2000. Logratio analysis and compositional distance. *Math. Geol.* 32 (3), 271–275.
- Aitchison, J., Egozcue, J., 2005. Compositional data analysis: where are we and where should we be heading? *Math. Geol.* 37, 829–850. <https://doi.org/10.1007/s11004-005-7383-7>.
- Aitchison, J., Greenacre, M., 2002. Biplots of compositional data. *J. Roy. Statist. Soc., C (Appl. Statist.)* 51 (4), 375–392.
- Albanese, S., 2008. Evaluation of the bioavailability of potentially harmful elements in urban soils through ammonium acetate-EDTA extraction: a case study in southern Italy. *Geochem. Explor. Environ. Anal.* 8, 49–57.

- Albanese, S., De Vivo, B., Lima, A., Cicchella, D., 2007. Geochemical background and baseline values of toxic elements in stream sediments of Campania region (Italy). *J. Geochem. Explor.* 93 (1), 21–34.
- Albanese, S., De Vivo, B., Lima, A., Cicchella, D., Civitillo, D., Cosenza, A., 2010. Geochemical baselines and risk assessment of the Bagnoli brownfield site coastal sea sediments (Naples, Italy). *J. Geochem. Explor.* 105, 19–33.
- Albanese, S., Fontaine, B., Chen, W., Lima, A., Cannatelli, C., Piccolo, A., et al. 2015a. Polycyclic aromatic hydrocarbons in the soils of a densely populated region and associated human health risks: The Campania plain (southern Italy) case study. *Environmental Geochemistry and Health*, 37, 1–20.
- Albanese, S., Iavazzo, P., Adamo, P., Lima, A., De Vivo, B., 2013. Assessment of the environmental conditions of the Sarno River Basin (South Italy): a stream sediment approach. *Environ. Geochem. Health* 35, 283–297.
- Albanese, S., Sadeghi, M., Lima, A., Cicchella, D., Dinelli, E., Valera, P., Falconi, M., Demetriades, A., De Vivo, B., The GEMAS project team, 2015b. GEMAS: Cobalt, Cr and Ni distribution in agricultural and grazing land soil of Europe. *J. Geochem. Explor.* 154, 81–93.
- Alves, Andréa A. R. et al. 2012. Comparison between GC-MS-SIM and GC-ECD for the determination of residues of organochlorine and organophosphorus pesticides in Brazilian citrus essential oils. *J. Braz. Chem. Soc.* [online]. 2012, vol.23, n.2, pp.306-314. ISSN 0103-5053. <http://dx.doi.org/10.1590/S0103-50532012000200017>. ATSDR (Agency for Toxic Substances and Disease Registry) 1995. Chlordane (CAS#12789-03-6). <http://www.atsdr.cdc.gov/toxfaqs/tfacts31.pdf> (Feb. 26, 2014).
- Ander, E.L., Johnson, C.C., Cave, M.R., Palumbo-Roe, B., Nathanail, C.P., Lark, R.M., 2013. Methodology for the determination of normal background concentrations of contaminants in English soil. *Sci. Total Environ.* 454–455, 604–618. <https://doi.org/10.1016/j.scitotenv.2013.03.005>.
- APAT (Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici), 2008. Criteri metodologici per l'applicazione dell'analisi assoluta di rischio ai siti contaminati. Revisione 2 APAT, Roma, 156 pp. <http://www.isprambiente.gov.it/files/temi/siti-contaminati-02marzo08.pdf>.
- APAT-ISS, 2006. Protocollo Operativo per la determinazione dei valori di fondo di metalli/ metalloidi nei suoli dei siti d'interesse nazionale. Revisione 0. Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici and Istituto Superiore di Sanita (in Italian).
- Aral, M., 2010. *Environmental Modeling and Health Risk Analysis (Acts/Risk)*. Springer, Dordrecht Heidelberg London New York. <http://dx.doi.org/10.1007/987-90-481-8608-2>.
- Aramendia, M. A.; Borau, V.; Lafont, F.; Marinas, A.; Marinas, J. M.; Moreno, J. M.; Urbano, F. J. 2007. Determination of herbicide residues in olive oil by gas chromatography–tandem mass spectrometry. *Food Chem.* 105 (2), 855–861.

- Arienzo, M., Adamo, P., Rosaria Bianco, M., Violante, P., 2001. Impact of land use and urban runoff on the contamination of the Sarno River Basin in southwestern Italy. *Water Air Soil Pollut.* 131, 349–366.
- Arienzo, M., Albanese, S., Lima, A., Cannatelli, C., Aliberti, F., Cicotti, F., et al. (2015). Assessment of the concentrations of polycyclic aromatic hydrocarbons and organochlorine pesticides in soils from the Sarno River basin, Italy, and ecotoxicological survey by *Daphnia magna*. *Environmental Monitoring and Assessment*, 187, 1-14.
- ATSDR (Agency for Toxic Substances and Disease Registry, USA), 2000. Toxicological Profile for Endosulfan. ATSDR, Atlanta, GA, USA.
- Ayuso, R.A., De Vivo, B., Rolandi, G., Seal II, R.R., Paone, A., 1998. Geochemical and isotopic (Nd-Pb-Sr-O) variations bearing on the genesis of volcanic rocks from Vesuvius. Italy. *J. of Volc. Geoth. Res.* 82, 53–78.
- Barber J.L., Sweetman, A.J., Van Wijk, D., Jones, K.C., 2005. Hexachlorobenzene in the global environment: emissions, levels, distribution, trends and processes. *Sci. Total Environ.* 349, 1-44.
- Barceló-Vidal, C., Martín-Fernández, J.A., 2016. The mathematics of compositional analysis. *Austrian J. Stat.* 45 (4), 57–71.
- Barkouch, Y., Sedki, A., Ponesu, A., 2007. A new approach for understanding lead transfer in agricultural soil. *Water, Air, Soil Pollut.* 186 (1–4), 3–13. <https://doi.org/10.1007/s11270-007-9450-9>.
- Bassot, J.P., 1997. Albitisations dans le Paléo-protérozoïque de l'Est du Sénégal: relations avec les minéralisations ferrifères de la rive gauche de la Falémé. *J Afr Earth Sci* 25:353–367. [French].
- Battisti, S., A. Caminiti, G. Ciotoli et al. 2013. 'A spatial, statistical approach to map the risk of milk contamination by beta-hexachlorocyclohexane in dairy farms'. *Geospatial Health* 8(1): 77–86.
- Bevilacqua, M., Braglia, M. 2002. Environmental efficiency analysis for ENI oil refineries. *Journal of Cleaner Production*, 10(1), 85–92.
- Bhupander, K., Gargi, G., Richa, G., Dev, P., Sanjay, K., Shekhar, S. C. 2012. Distribution, composition profiles and source identification of polycyclic aromatic hydrocarbons in roadside soil of Delhi, India. *Journal of Environmental Earth Sciences*, 2(1), 10–22.
- Bianconi D., De Paolis M.R., Agnello M.C., Lippi D, Pietrini F, Zacchini M, Polcaro C, Donati E, Paris P, Spina S, Massacci A. 2010. Field-scale rhizoremediation of contaminated soil with hexachlorocyclohexane (HCH) isomers: the potential of poplars for environmental restoration and economical sustainability. In: Golubev IA (ed) *Handbook of phytoremediation*. Nova Science. Publishers Inc, Hauppauge, pp. 783-794.
- Bidleman, T.F., Leone, A.D., 2004. Soil-air exchange of organochlorine pesticides in the Southern United States (US). *Environmental Pollution* 128, 49–57.

- Bixiong, Y., Zhihuan, Z., Ting, M. 2006. Pollution sources identification of polycyclic aromatic hydrocarbons of soils in Tianjin area, China. *Chemosphere*, 64(4), 525–534.
- Björklund, A., Gustavsson, N., 1987. Visualization of geochemical data on maps: new options. *J. Geochem. Explor.* 29, 89–103.
- Bollen, K.A., 1987. Outliers and improper solutions: a confirmatory factor analysis example. *Sociol. Methods Res.* 15, 375–384.
- Bonardi, G., D'Argenio, D., Perrone, V., 1998. Carta geologica dell'Appennino meridionale. *Mem. Soc. Geol. Ital.* 41.
- Boni, M., Rollinson, G., Mondillo, N., Balassone, G., Santoro, L., 2013. Quantitative mineralogical characterization of karst bauxite deposits in the southern Apeninnes. Italy. *Econ. Geol.* 108, 813–833.
- British Geological Survey (BGS) Technical Report WP/95/14. Price, J., Velbel, M., 2003. Chemical weathering indices applied to weathering profiles developed on heterogeneous felsic metamorphic parent rocks. *Chem. Geol.* 202, 4397–4416. <http://dx.doi.org/10.1016/j.chemgeo.2002.11.001>.
- Buat-Ménard, P., Chesselet, R., 1979. Variable influence of the atmospheric flux on the trace metal chemistry of oceanic suspended matter. *Earth Planet. Sci. Lett.* 42, 399–411.
- Buccianti, A., Lima, A., Albanese, S., Cannatelli, C., Esposito, R., De Vivo, B., 2015. Exploring topsoil geochemistry from the CoDA (Compositional Data Analysis) perspective: the multi-element data archive of the Campania Region (Southern Italy). *J. Geochem. Explor.* 159, 302–316.
- Buccianti, A., Lima, A., Albanese, S., De Vivo, B., 2018. Measuring the change under Compositional Data Analysis (CoDA): insight on geochemical system dynamics. *J. Geochem. Explor.* 189, 100–108. <https://doi.org/10.1016/j.gexplo.2017.05.006>.
- Buccianti, A., Magli, R., 2011. Metric concepts and implications in describing compositional changes for world river's water chemistry. *Comput. Geosci.* 37 (5), 670–676.
- Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V., 2006. *Compositional Data Analysis in the Geosciences: From Theory to Practice*.
- Buccianti, A., Pawlowsky Glahn, V., Egoscue, J.J., 2014. Variation diagrams to statistically model the behaviour of geochemical variables: theory and applications. *J. Hydrogeol.* 519, 988–998.
- Bucheli, T. D., Blum, F., Desaulles, A., Gustaffson, O. 2004. Polycyclic aromatic hydrocarbons, black carbon, and molecular markers in soils of Switzerland. *Chemosphere*, 56, 1061–1076.
- Budzinski, H., Jones, I., Bellocq, J., Pierard, C., Garrigues, P. 1997. Evaluation of sediment contamination by polycyclic aromatic hydrocarbons in the Gironde estuary. *Marine Chemistry*, 58, 85–97.

- Bundschuh, J., Litter, M.I., Parvez, F., Román-Ross, G., Nicolli, H.B., Jean, J.-S., Liu, C.-W., López, D., Armienta, M.A., Guilherme, L.R.G., Cuevas, A.G., Cornejo, L., Cumbal, L., Toujaguez, R. 2012. One century of arsenic exposure in Latin America: a review of history and occurrence from 14 countries. *The Science of the Total Environment*, 429: 2–35. <https://doi.org/10.1016/j.scitotenv.2011.06.024>.
- Bureau Veritas Minerals, 2017. AQ250 (Ultra Trace Geochemical Aqua Regia Digestion), vols. 1–2.
- Caboche, J. 2009. Validation d'un test de mesure de bio-accessibilité - Application à 4 éléments traces métalliques dans les sols : As, Cd, Pb et Sb. Laboratoire unité de recherche animale et fonctionnalités des produits animaux (URAFPA), Nancy, Institut national polytechnique de Lorraine [French].
- Calvelo Pereira, R., Camps-Arbestain, M., Rodríguez Garrido, B., Macías, F., Monterroso, C., 2006. Behaviour of α -, β -, γ - and δ -hexachlorocyclohexane in the soil-plant system of a contaminated site. *Environmental Pollution*, 144, 210-217.
- Campbell, M., Norstrom, R.J., Hobson KA, Muir DCG, Backus S, Fisk AT. 2005. Mercury and other trace elements in a pelagic arctic marine food web (Northwater Polynya, Baffin Bay). *Sci Total Environ* 351–352:247–263.
- Cantu-Soto, E. U., Meza-Montenegro, M. M., Valenzuela-Quintanar, A. I., Felix-Fuentes, A., Grajeda-Cota, P., Balderas-Cortes, J. J., et al. (2011). Residues of organochlorine pesticides in soils from the Southern Sonora, Mexico. *Bulletin of Environmental Contamination and Toxicology*, 87(5), 556–560.
- Carranza, E.J.M., 2011. Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values. *J. Geochem. Explor.* 110, 167–185.
- Cave, M.R., Johnson, C.C., Ander, E.L., Palumbo-Roe, B., 2012. Methodology for the determination of normal background contaminant concentrations in English soils. In: British Geological Survey Commissioned Report, CR/12/003, (41 pp.). <http://nora.nerc.ac.uk/19959/>.
- CCME Canadian Council of Ministers of the Environment. 2007. Soil Quality Index 1.0: Technical Report. In: Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg.
- CCME. 1992. National classification system for contaminated sites. Report # CCME EPC-CS39E. Canadian Council of Ministers of the Environment, Winnipeg.
- Cerrillo I., Granada A., Lopez-Espinosa M.J., Olmos B., Jimenez M., Cano A., Olea N., Olea-Serrano MF. (2005). Endosulfan and its metabolites in fertile women, placenta, cord blood, and human milk. *Environ Res.* 98, 233–239.
- Chayes, F., 1960. On correlation between variables of constant sum. *J. Geophys. Res.* 65 (12), 4185–4193.

- Chen, L., Ran, Y., Xing, B., Mai, B., He, J., Wei, X., et al. 2005. Contents and sources of polycyclic aromatic hydrocarbons and organochlorine pesticides in vegetable soils of Guangzhou, China. *Chemosphere*, 60, 879–890.
- Cheng, Q., 1999a. Multifractality and spatial statistics. *Comput. Geosci.* 25, 949–961.
- Cheng, Q., 1999b. Spatial and scaling modelling for geochemical anomaly separation. *J. Geochem. Explor.* 65, 175–194.
- Cheng, Q., 1999c. In: Lippard, S.J., Naess, A., Sinding-Larsen, R. (Eds.), *Multifractal interpolation*. Proc. Ann. Conf. International Association for Mathematical Geology, Trondheim, Norway 6–11 August 1999. Vol. 1. Norwegian University of Science and Technology, Trondheim, pp. 245–250.
- Cheng, Q., 2000a. Interpolation by means of multifractal, kriging and moving average techniques. *GeoCanada 2000*. In: Proc. GAC/MAC Meeting, Calgary, AB, Canada [CDROM], 29 May–2 June 2000. Geol. Assoc. Can., St. John's, NF, Canada.
- Cheng, Q., 2000b. *GeoData Analysis System (GeoDAS) for Mineral Exploration: Unpublished User's Guide and Exercise Manual*. Material for the Training Workshop on GeoDAS held at York University.
- Cheng, Q., 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. *Ore Geol. Rev.* 32, 314–324.
- Cheng, Q., 2008. Modeling local scaling properties for multiscale mapping. *Vadose Zone J.* 7, 525–532.
- Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. *J. Geochem. Explor.* 51, 109–130.
- Cheng, Q., Bonham-Carter, G.F., Raines, G.L., 2001. GeoDAS: a new GIS system for spatial analysis of geochemical data sets for mineral exploration and environmental assessment. In: *The 20th Intern. Geochem. Explor. Symposium (IGES)*. Santiago de Chile. Vol. 6/5–10/5. pp. 42–43.
- Cheng, Q., Xia, Q., Li, W., Zhang, S., Chen, Z., Zuo, R., Wang, W., 2010. Density/area power-law models for separating multi-scale anomalies of ore and toxic elements in stream sediments in Gejiu mineral district, Yunnan Province, China. *Biogeosciences* 7, 3019–3025.
- Chester, R., Stoner, J.H., 1973. Pb in particulates from the lower atmosphere of the eastern Atlantic. *Nature* 245, 27–28.
- Cicchella, D., De Vivo, B., Lima, A., 2005. Background and baseline concentration values of elements harmful to human health in the volcanic soils of the metropolitan and provincial area of Napoli (Italy). *Geochem. Explor. Environ. Anal.* 5, 29–40.
- Cicchella, D., Giaccio, L., Dinelli, E., Albanese, S., Lima, A., Zuzolo, D., Valera, P., De Vivo, B., 2015. GEMAS: Spatial distribution of chemical elements in agricultural and grazing land soil of Italy. *J. Geochem. Explor.* 154, 129–142.

- Cicchella, D., Giaccio, L., Lima, A., Albanese, S., Cosenza, A., Civitillo, D., De Vivo, B., 2014. Assessment of the topsoil heavy metals pollution in the Sarno River Basin, south Italy. *Environ. Earth Sci.* 71, 5129–5143.
- Cicchella, D., Hoogewerff, J., Albanese, S., Adamo, P., Lima, A., Taiani, M.V.E., De Vivo, B., 2016. Distribution of toxic elements and transfer from the environment to human traced by using lead isotopes. A case study in the Sarno River basin, south Italy. *Environ. Geochem. Health* 38, 619–637.
- Clarkson, T.W., Magos L. 2006. The toxicology of mercury and its chemical compounds. *Crit Rev Toxicol.* (in press).
- CoDaWork'11: 4th International Workshop on Compositional Data Analysis. SantFeliu de Guíxols.
- Comas-Cufí, M., Thió-Henestrosa, S., 2011. CoDaPack 2.0: a stand-alone, multi-platform compositional software. In: Egozcue, J.J., Tolosana-Delgado, R., Ortego, M.I. (Eds.), *CoDaWork'11: 4th International Workshop on Compositional Data Analysis*. SantFeliu de Guíxols.
- Conticelli, S., Marchionni, S., Rosa, D., Giordano, G., Boari, E., Avanzinelli, R., 2009. Shoshonite and sub-alkaline magmas from an ultrapotassic volcano: Sr–Nd–Pb isotope data on the Roccamonfina volcanic rocks, Roman Magmatic Province, Southern Italy. *Contrib. Mineral. Petrol.* 157 (1), 41–63.
- Coordination de l'information sur l'Environnement Land Cover (Corine Land Cover, Italy) 2012. <http://www.pcn.minambiente.it/geoportal/catalog/search/resource/details.page>.
- Corona, P., Barbati, A., Tomao, A., Bertani, R., Valentini, R., Marchetti, M., Perugini, L., 2012. Land use inventory as framework for environmental accounting: an application in Italy. *IForest-Biogeosciences and Forestry* 5 (4), 204. <http://dx.doi.org/10.3832/ifor0625-005>.
- Costantini, E.A.C., Dazzi, C., 2013. *The Soils of Italy*. Springer (354 pp.).
- Counter, S. A., and Buchanan, L. H. 2004. Mercury exposure in children: a review. *Toxicology and Applied Pharmacology*, 198,209–230.
- Courtney K.D. 1979. Hexachlorobenzene (HCB): a review. *Environ Res.* 20, 225–66.
- De Pippo, T., Donadio, C., Guida, M., et al., 2006. *Environ. Sci. Pollut. Res. Int.* 13, 184. <http://dx.doi.org/10.1065/espr2005.08.287>.
- De Vivo, B., Ayuso, R.A., Belkin, H.E., Fedele, L., Lima, A., Rolandi, G., Somma, R., Webster, J.D., 2003. Chemistry, fluid/melt inclusions and isotopic data of lavas, tephra and nodules from 25 ka to 1944 AD of the Mt. Somma-Vesuvius volcanic activity. In: *Mt. Somma-Vesuvius Geochemical Archive*. Dipartimento di Geofisica e Vulcanologia, Università di Napoli Federico II, Open File Report 1–2003, (143 pp).
- De Vivo, B., Lima, A., Albanese, S., Cicchella, D., Rezza, C., Civitillo, D., Minolfi, G., Zuzolo, D., 2016. *Atlante geochimico-ambientale dei suoli della Campania (Environmental Geochemical Atlas of Campania Soils)*. Aracne Editrice, Roma. 364 pp. (ISBN 978-88-548-9744-1. in Italian).

- De Vivo, B., Petrosino, P., Lima, A., Rolandi, G., Belkin, H.E., 2010. Research progress in volcanology in Neapolitan area, Southern Italy: a review and alternative views. *Mineral. Petrol.* 99, 1–28. <https://doi.org/10.1007/s00710.009.0098.6>.
- De Vivo, B., Rolandi, G., Gans, P.B., Calvert, A., Bohrsen, W.A., Spera, F.J., Belkin, H.E., 2001. New constraints on the pyroclastic eruptive history of the Campanian volcanic Plain (Italy). *Mineral. Petrol.* 73, 47–65.
- DEA. 2010. Framework for the Management of Contaminated Land. Republic of South Africa, Department of Environmental Affairs. (also available at <http://sawic.environment.gov.za/documents/562.pdf>).
- Dearth, M.A., Hites, R.A., 1991. Complete analysis of technical chlordane using negative ionization mass spectrometry. *Environ. Sci. Technol.* 25, 245–254.
- DEFRA (Department for Environment, Food and Rural Affairs), 2011. Guidelines for Environmental Risk Assessment and Management: Green Leaves III. Defra and the Collaborative Centre of Excellence in Understanding and Managing Natural and Environmental Risks. Cranfield University, UK.
- Di Gennaro, A., Aronne, G., De Mascellis, R., Vingiani, S., Sarnataro, M., Abalsamo, P., Cona, F., Vitelli, L., Arpaia, G., 2002. I sistemi di terre della Campania. In: *Monografia e Carta*. 1. pp. 250.000. <http://hdl.handle.net/11588/179891>.
- DoE (Department of Environment), 1995. A Guide to Risk Assessment and Risk Management for Environmental Protection. HSMO, London.
- Dogliani, C., Flores, G. 1997. Italy. In E. M. Moores R. W. Fairbridge (Eds.), *Encyclopedia of European and Asian regional geology* (pp. 414–435). New York: Chapman Hall.
- Dong, T., Lee, B. K. 2009. Characteristics, toxicity, and source apportionment of polycyclic aromatic hydrocarbons (PAHs) in road dust of Ulsan, Korea. *Chemosphere*, 74, 1245–1253.
- Donnarumma L., Pompei, V., Faraci, A., Conte, E. 2009. Dieldrin uptake by vegetable crops grown in contaminated soils, *J. Environ. Sci. Health, Part B.* 44, 449–454.
- DPCM. 1990. Decreto Presidente del Consiglio dei Ministri (DPCM), 30 Novembre 1990 (Italian).
- Ducci, D., Albanese, S., Boccia, L., Celentano, E., Cervelli, E., Corniello, A., Crispo, A., De Vivo, B., Iodice, P., Langella, C., Lima, A., Manno, M., Palladino, M., Pindozzi, S., Rigillo, M., Romano, N., Sellerino, M., Senatore, A., Speranza, G., Fiorentino, N., Fagnano, M., 2017. An integrated approach for the environmental characterization of a wide potentially contaminated area in southern Italy. *Int. J. Environ. Res. Publ. Health* 14, 1–23.
- Ducci, D., Tranfaglia, G., 2005. L'impatto dei cambiamenti climatici sulle risorse idriche sotteranee in Campania. *Boll. Ordine dei Geologi Della Campania* 1-4, 13–21 (in Italian).
- Duce, R.A., Hoffmann, G.L., Zoller, W.H., 1975. Atmospheric trace metals at remote northern and southern hemisphere sites: pollution or natural? *Science* 187, 59–61.

- Durn, G., Slovenec, D., Čović, M., 2001. Distribution of iron and manganese in Terra Rossa from Istria and its genetic implications. *Geologica Croatica* 54, 27–36.
- Eckart, G., Young, G., 1936. The approximation of one matrix by another of lower rank. *Psychometrika* 1 (3), 211–218.
- EEA. 2014. Progress in management of contaminated sites. European Environment Agency. (also available at <https://www.eea.europa.eu/data-and-maps/indicators/progress-in-management-of-contaminated-sites/progress-in-management-ofcontaminated-1>).
- Egli, M., Mirabella, A., Sartori, G., 2008. The role of climate and vegetation in weathering and clay mineral formation in late Quaternary soils of the Swiss and Italian Alps. *Geomorphology* 102, 307–324. <https://doi.org/10.1016/j.geomorph.2008.04.001>.
- Egozcue, J.J., 2009. Reply to “On the Harker variation diagrams. by J. A. Cortés. *Math Geosci.* 41 (7), 829–834.
- Egozcue, J.J., Pawlowsky-Glahn, V., 2005. Groups of parts and their balances in compositional data analysis. *Math. Geol.* 37 (7), 795–828.
- Egozcue, J.J., Pawlowsky-Glahn, V., 2006. Simplicial geometry for compositional data. In: Buccianti, A.A., Mateu-Figueras, G., Pawlowsky-Glahn, V. (Eds.), *Compositional Data Analysis in the Geosciences. From Theory to Practice*, vol. 264. Geological Society, London, Special Publications, pp. 145–159.
- Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-figueras, G., Barcelo-vidal, C., 2003. Isometric logratio transformations for compositional data analysis. *Math. Geol.* 35 (3), 279–300.
- Endosulfan Preliminary Dossier, 2003. Umweltbundesamt, Berlin, pp 55.
- Environmental Protection Agency (EPA). 1984. Health effects assessment for polycyclic aromatic hydrocarbons (PAH). EPA 540/I-86-013. Environmental Criteria and Assessment Office, Cincinnati, OH.
- ESRI (Environmental Systems Research Institute), 2012. ArcGIS Desktop: Release 10. Redlands, CA.
- Estellano V.H., Pozo, K., Harner, T., Corsolini, S., Focardi, S., (2012). Using PUF disk passive samplers to simultaneously measure air concentrations of persistent organic pollutants (POPs) across the Tuscany. Region, Italy. *Atmos Pollut Res.* 3, 88-94.
- Eurostat. 2014. http://ec.europa.eu/eurostat/statistics-explained/index.php/Pesticide_sales_statistics.explained/index.php/Pesticide_sales_statistics#Data_sources_and_availability
- Fabian, C., Reimann, C., Fabian, K., Birke, M., Baritz, R., Haslinger, E., The GEMAS Project Team, 2014. GEMAS: spatial distribution of the pH European agricultural and grazing land soil. *Appl. Geochem.* 48, 207–216. <http://dx.doi.org/10.1016/j.apgeochem.2014.07.017>.
- Fabian, K., Reimann, C., Caritat, P. de, 2017. Quantifying diffuse contamination: method and application to Pb in soil. *Environ. Sci. Technol.* 51, 6719–6726.

- Falandysz, J., Brudnowska, B., Kawano, M., Wakimoto, T., 2001. Polychlorinated biphenyls and organochlorine pesticides in soils from the southern part of Poland. *Arch. Environ. Contam. Toxicol.* 40, 173-178.
- Fan, M.A., and G.V. Alexeeff, 1999. Public Health Goal for Endrin in Drinking Water. Office of Environmental Health and Hazard Assessment. Environmental Protection Agency, California, pp: 5-6.
- Fang, G.-C., Chang, K.-C., Lu, C., Bai, H. 2004. Estimation of PAHs dry deposition and BaP toxic equivalency factors (TEFs) study at Urban, Industry Park and rural sampling sites in central Taiwan, Taichung. *Chemosphere*, 55(6), 787–796. <https://doi.org/10.1016/j.chemosphere.2003.12.012>.
- Fang, Y., Nie, Z., Die, Q., Tian, Y., Liu, F., He, J., Huang, Q., 2017. Organochlorine pesticides in soil, air, and vegetation at and around a contaminated site in southwestern China: concentration, transmission, and risk evaluation. *Chemosphere*. 178, 340-349. <https://doi.org/10.1016/j.chemosphere.2017.02.151>.
- Filzmoser, P., Hron, K., 2008. Outlier detection for compositional data using robust methods. *Math. Geosci.* 40 (3), 233–248.
- Filzmoser, P., Hron, K., 2011. Robust statistical analysis. In: Pawlowsky-Glahn, V., Buccianti, A. (Eds.), *Compositional Data Analysis. Theory and Applications*. John Wiley Sons, Chichester, UK, pp. 59–72.
- Filzmoser, P., Hron, K., Reimann, C., 2009a. Principal component analysis for compositional data with outliers. *Environmetrics* 20 (6), 621–632.
- Filzmoser, P., Hron, K., Reimann, C., 2009b. Univariate statistical analysis of environmental (compositional) data - problems and possibilities. *Sci. Total Environ.* 407, 6100–6108.
- Filzmoser, P., Hron, K., Tolosana-Delgado, R., 2012. Interpretation of multivariate outliers for compositional data. *Comput. Geosci.* 39, 77–85.
- Gabriel, K.R., 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58 (3), 453.
- Giannetti, B., Luhr, J.F., 1983. The white trachytic tuff of Roccamonfina volcano (roman region, Italy). *Contrib. Mineral. Petrol.* 84, 235–252.
- Goldberg, E.D., 1972. Baseline studies of pollutants in the marine environment and research recommendations. In: *The International Decade of Ocean Exploration (IDOE) Baseline Conference*, 24–26 May 1972. New York.
- Gosar, M., Šajn, R., Biester, H., 2006. Binding of mercury in soils and attic dust in the Idrija mercury mine area (Slovenia). *Sci. Total Environ.* 369, 150–162.
- Guillén, M.T., Delgado, J., Albanese, S., Nieto, J.M., Lima, A., De Vivo, B., 2011. Environmental geochemical mapping of Huelva municipality soils (SW Spain) as a tool to determine background

- and baseline values. *J. Geochem. Explor.* 109 (1–3), 59–69. <https://doi.org/10.1016/j.gexplo.2011.03.003>.
- Hakanson, L., 1980. An ecological risk index for aquatic pollution control. A sedimentological approach. *Water Res.* 14 (8), 975–1001. [https://doi.org/10.1016/0043-1354\(80\)90143-8](https://doi.org/10.1016/0043-1354(80)90143-8).
- Han, J., Kamber, M., 2001. *Data Mining: Concepts and Techniques*. Morgan-Kaufmann Academic Press, San Francisco.
- Hawkes, H.E., Webb, J.S., 1962. *Geochemistry in Mineral Exploration*. Harper, New York, pp. 415.
- Health Canada, 2010a. Federal Contaminated Site Risk Assessment in Canada. Part I: Guidance on Human Health Preliminary Quantitative Risk Assessment (PQRA). Version 2.0. Health Canada, Ottawa.
- Health Canada, 2010b. Federal Contaminated Site Risk Assessment in Canada, Part II: Health Canada Toxicological Reference Values (TRVs) and Chemical-Specific Factors. Version 2.0. Health Canada, Ottawa.
- Hron, K., Templ, M., Filzmoser, P., 2010. Imputation of missing values for compositional data using classical and robust methods. *Comput. Stat. Data Anal.* 54 (12), 3095–3107.
- Hwang, H. M., Wade, T. L., Sericano, J. L. 2003. Concentrations and source characterization of polycyclic aromatic hydrocarbons in pine needles from Korea, Mexico, and United States. *Atmospheric Environment*, 37, 2259–2267.
- IARC (International Agency for Research on Cancer). 1983. IARC monographs on the evaluation of the carcinogenic risk of chemicals to human. Polynuclear aromatic compounds, Part I, chemical, environmental, and experimental data. Geneva: World Health Organization.
- Ishikawa, T., Ikegaki, Y. 1980. Control of mercury pollution in Japan and the Minamata Bay clean-up. *Journal (Water Pollution Control Federation)*, 52 (5), 1013– 1018.
- Islam, M. N., Park, M., Jo, Y. T., Nguyen, X. P., Park, S. S., Chung, S.-Y., 2017. Distribution, sources, and toxicity assessment of polycyclic aromatic hydrocarbons in surface soils of the Gwangju City, Korea. *Journal of Geochemical Exploration*, 180, 52–60. <https://doi.org/10.1016/j.gexplo.2017.06.009>.
- ISPRA 2014. Audizione dell'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) presso la Commissione Agricoltura, congiuntamente con la Commissione Ambiente, della Camera sul consumo di suolo, Audizione, Roma, 27/2/2014 [Italian].
- ISPRA Istituto SperlaPAmbientale. (2012). Qualità dell'ambiente urbano. VIII Rapporto. ISPRA-ARPA-APPA, roma (Italian).
- ISPRA, 2014b. Il consumo di suolo in Italia. Retrieved from <http://www.isprambiente.gov.it/it/pubblicazioni/rapporti/il-consumo-di-suolo-in-italia> [in Italian].

- ISTAT, 2013. Resident Population on 1 January 2013 Available Via Dialog. <http://www.istat.it/it/campania/dati> , Accessed date: 25 April 2016.
- ISTAT, 2016. Resident population on 1st January, 2017. <https://www.istat.it/en/population-and-households>. .
- Jarauta-Bragulat, E., Hervada-Sala, C., Egozcue, J.J., 2015. Air quality index revisited from a compositional point of view. *Math. Geosci.* <https://doi.org/10.1007/s11004-015-9599-5>.
- Jayaraj, R., Megha, P., Sreedev, P. 2016. Organochlorine pesticides, their toxic effects on living organisms and their fate in the environment. *Interdisciplinary Toxicology*. 9(3-4): 90–100, doi:10.1515/intox-2016-0012.
- Jennings, A., Li, Z., 2014. Scope of the worldwide effort to regulate pesticide contamination in surface soils. *J. Environ. Manag.* 146, 420-443.
- Ji, Y., Feng, Y., Wu, J., Zhu, T., Bai, Z., Duan, C., 2008. Using Geoaccumulation index to study source profiles of soil dust in China. *J. Environ. Sci.* 20 (5), 571–578. [https://doi.org/10.1016/S1001-0742\(08\)62096-3](https://doi.org/10.1016/S1001-0742(08)62096-3).
- Jia, H., Liu, L., Sun, Y., Sun, B., Wang, D., Su, Y., Kannan, K., Li, Y.-F., 2010. Monitoring and modelling endosulfan in Chinese surface soil. *Environ. Sci. Technol.* 44, 9279–9284.
- Jiang, Y.F., Wang, X.T., Jia, Y., Wang, F., Wu, M.H., Sheng, G.Y., Fu, J.M., 2009. Occurrence, distribution and possible sources of organochlorine pesticides in agricultural soil of Shanghai, China. *J. Hazard Matter* 170, 989-997.
- Joron, J.L., Metrich, N., Rosi, M., Santocroce, R., Sbrana, A., 1987. Chemistry and petrography. In: Santacorce, R. (Ed.), *Somma-Vesuvius*. CNR Quad. Ric. Sci. Vol. 114. pp. 105–174.
- Kabata-Pendias, A., 2011. Trace Elements of Soils and Plants, fourth ed. CRC Press, Taylor Francis Group, LLC, USA, pp. 28–534.
- Katsoyiannis, A., Terzi, E., Cai, Q.-Y. 2007. On the use of PAH molecular diagnostic ratios in sewage sludge for the understanding of the PAH sources. Is this use appropriate? *Chemosphere*, 69, 1337–1339.
- Kempton, R.A., 1984. The use of biplots in interpreting variety by environment interactions. 103 (01), 123–135.
- Khalili, N. R., Scheff, P. A., Holsen, T. M. 1995. PAH source fingerprints for coke ovens, diesel and gasoline engine highway tunnels and wood combustion emissions. *Atmospheric Environment*, 29, 533–542.
- Kim, K. H., Kabir, E. and Jahan, S. A. 2017. Exposure to pesticides and the associated human health effects. *Sci. Total Environ.* 575, 525–535.
- Kozak, K., Ruman, M., Kosek, K., Karasiński, G., Stachnik, L., Żaneta Polkowska, Z. 2017. Impact of volcanic eruptions on the occurrence of PAHs compounds in the aquatic ecosystem of the

- Southern Part of West Spitsbergen (Hornsund Fjord, Svalbard). *Water*, 9(1), 42. <https://doi.org/10.3390/w9010042>.
- Kürzl, H., 1988. Exploratory data analysis: recent advances for the interpretation of geochemical data. *J. Geochem. Explor.* 30, 309–322.
- Larsen, R. K., III, Baker, J. E. 2003. Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: A comparison of three methods. *Environmental Science and Technology*, 37, 1873–1881.
- Legislative Decree 152/2006 Decreto Legislativo 3 aprile 2006, n. 152, “Norme in materia ambientale”. *Gazzetta Ufficiale* n. 88 14-4-2006, Suppl. Ord. n. 96. <http://www.camera.it/parlam/leggi/deleghe/06152dl.htm>.
- Lepeltier, C., 1969. A simplified statistical treatment of geochemical data by graphical representation. *Econ. Geol.* 64, 538–550.
- Li X.H., Wang, W., Wang, J., Cao, X.L., Wang, X.F., Liu, J.C., 2008. Contamination of soils with organochlorine pesticides in urban parks in Beijing, China. *Chemosphere*. 70, 1660–1668.
- Li, W. H., Tian, Y. Z., Shi, G. L., Guo, C. S., Li, X., Feng, Y. C. 2012. Concentrations and sources of PAHs in surface sediments of the Fenhe reservoir and watershed, China. *Ecotoxicology and Environmental Safety*, 75, 198–206. <https://doi.org/10.1016/j.ecoenv.2011.08.021>.
- Lim, H.S., Lee, J.S., Chon, H.T., Sager, M., 2008. Heavy metal contamination and health risk assessment in the vicinity of the abandoned Songcheon Au–Ag mine in Korea. *J. Geochem. Explor.* 96 (2–3), 223–230. <https://doi.org/10.1016/j.gexplo.2007.04.008>.
- Lima, A., Albanese, S., De Vivo, B., 2005. Geochemical baselines for the radioelements K, U and Th in the Campania region, Italy: a comparison of stream-sediment geochemistry and gamma-ray surveys. *Appl. Geochem.* 20, 611–625.
- Lima, A., De Vivo, B., Cicchella, D., Cortini, M., Albanese, S., 2003a. Multifractal IDW interpolation and fractal filtering method in environmental studies: an application on regional stream sediments of Campania Region (Italy). *Appl. Geochem.* 18 (12), 1853–1865. [https://doi.org/10.1016/S08832927\(03\)00083-0](https://doi.org/10.1016/S08832927(03)00083-0).
- Lima, A., Danyushevsky, L.V., De Vivo, B., Fedele, L., 2003b. A model for the evolution of the Mt. Somma-Vesuvius magmatic system based on fluid and melt inclusion investigations. In: De Vivo, B., Bodnar, R.J. (Eds.), *Melt Inclusions in Volcanic Systems: Methods, Applications and Problems*. Series: Development in Volcanology Elsevier, Amsterdam 272 pp.
- Loska, K., Cebula, J., Pelczar, J., Wiechula, D., Kwapilinski, J., 1997. Use of enrichment and contamination factors together with geoaccumulation indexes to evaluate the content of Cd, Cu, and Ni in the Rybnik water reservoir in Poland. *Water Air Soil Pollut.* 93, 347–365.
- Luo, Y., Wu, L., Liu, L., Han, C., Li, Z. 2009. Heavy Metal Contamination and Remediation in Asian Agricultural Land. p. 9. Paper presented at MARCO Symposium, 2009, Japan.

- Mackay, D., Shiu, W.Y., Ma, K.C., 1992. Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Compounds, Lewis Publishers, Chelsea, MI.
- Maliszewska-Kordybach, B., Smreczak, B., Klimkowicz-Pawlas, A. 2009. Concentrations, sources, and spatial distribution of individual polycyclic aromatic hydrocarbons (PAHs) in agricultural soils in the Eastern part of the EU: Poland as a case study. *Science of the Total Environment*, 407(12), 3746–3753. <https://doi.org/10.1016/j.scitotenv.2009.01.010>.
- Mandelbrot, B.B., 1983. *The Fractal Geometry of Nature*. Freeman, San Francisco, pp. 468.
- Manz, M., Wenz, K.D., Dietze, U., 2001. Persistent organic pollutants in agricultural soils of central Germany. *Sci. Total Environ.* 277, 187–198.
- Marfisi, G., Di Stefano, C. 2016. The evidence of toxic wastes dumping in Campania, Italy. *Critical Reviews in Oncology/Hematology*, 105, 84–91. <https://doi.org/10.1016/j.critrevonc.2016.05.007>.
- Maronna, R., Martin, R., Yohai, V., 2006. *Robust Statistics: Theory and Methods*. John Wiley, pp. 436 (ISBN: 978-0-470-01092-1).
- Martin, J.M., Whitfield, M., 1983. The significance of the river input of chemical elements to the ocean. In: Wong, C.S., Boyle, E., Bruland, K.W., Burton, J.D., Goldberg, E. (Eds.), *Trace Metals in Sea Water*. Plenum, New York, pp. 265–296.
- Masclet, P., Bresson, M. A., Mouvrier, G. 1987. Polycyclic aromatic hydrocarbons emitted by power stations, and influence of combustion conditions. *Fuel*, 66, 556–562.
- Mateu-Figueras, G., Daunis i Estadella, J., 2008. Compositional Amalgamations and Balances: A Critical Approach. CODAWORK'08. Available via Dialog. <http://hdl.handle.net/10256/738>.
- Mazza, A., Piscitelli, P., Neglia, C., Della Rosa, G., Iannuzzi, L. 2015. Illegal dumping of toxic waste and its effect on human health in Campania, Italy. *International Journal of Environmental Research and Public Health*, 12, 6818–6831. <https://doi.org/10.3390/ijerph120606818>.
- McKinley, J.M., Hron, K., Grunsky, E.C., Reimann, C., Caritat, P. de, Filzmoser, P., van den Boogaart, K.G., Tolosana-Delgado, R., 2016. The single component geochemical map: fact or fiction? *J. Geochem. Explor.* 162, 16–28.
- Means, J. C., Wood, S. G., Hassett, J. J., Banwart, W. L. 1980. Sorption of polynuclear aromatic hydrocarbons by sediments and soils. *Environmental Science and Technology*, 14(12), 1524–1528.
- Menzie, C. A., Potocki, B. B., Santodonato, J. 1992. Exposure to carcinogenic PAHs in the environment. *Environmental Science and Technology*, 26(7), 1278–1283.
- Metcalf, R.L. 1995. *Insect Control Technology*. In: Kroschwitz, J. and Howe-Grant, M. (eds.). *Kirk-Othmer Encyclopedia of Chemical Technology*. Volume 14. New York, NY: John Wiley and Sons, Inc. pp. 524-602.
- Miesch Programs, 1990. G-RFAC. Grand Junction, Colorado, USA.

- Miko, S., Durn, G., Prohić, E., 1999. Evaluation of terra rossa geochemical baselines from Croatian karst regions. *J. Geochem. Explor.* 66, 173–182.
- Minolfi, G., Albanese, S., Lima, A., Tarvainen, T., Fortelli, A., De Vivo, B., 2018a. A regional approach to the environmental risk assessment - human health risk assessment: case study in the Campania region. *J. Geochem. Explor.* 184 (Part B), 400–416.
- Minolfi, G., Petrik, A., Albanese, S., Lima, A., Rezza, C., De Vivo, B., 2018b. Lead, Cu and Zn distributions in topsoils of the Campania region, Italy. *Spec. Issue Geochem.: Explor. Environ. Anal.* <https://doi.org/10.1144/geochem2017-074>.
- Mishra, K., Sharma, R.C., Kumar, S., 2012. Contamination levels and spatial distribution of organochlorine pesticides in soils from India. *Ecotoxicol. Environ. Saf.* 76, 215–225.
- Moeckel, C., Macleod, M., Hungerbühler, K., Jones, K.C., 2008. Measurement and modelling of diel variability of polybrominated diphenyl ethers and Chlordanes in air. *Environmental Science Technology* 42, 3219–3225.
- Mondillo, N., Balassone, G., Boni, M., Rollinson, G., 2011. Karst bauxites in the Campania Apennines (southern Italy): a new approach. *Period. Mineral.* 80 (3), 407–432.
- Morillo, E., Romero, A. S., Madrid, L., Villaverde, J., Maqueda, C. 2008. Characterization and sources of PAHs and potentially toxic metals in urban environments of Seville (Southern Spain). *Water Air Soil Pollution*, 187, 41–51.
- Morillo, E., Romero, A. S., Maqueda, C., Madrid, L., Ajmone-Marsan, F., Grcman, H., et al. 2007. Soil pollution by PAHs in urban soils: A comparison of three European cities. *Journal of Environmental Monitoring*, 9, 1001–1008.
- Müller, G., 1969. Index of Geoaccumulation in sediments of the Rhine river. *Geo Journal* 2, 108–118.
- Müller, G., 1981. The heavy metal pollution of the sediments of Neckars and its tributary: a stock taking. *Chem. Ztg.* 105, 157–164.
- Nadal, M., Schuhmacher, M., Domingo, J. L. 2004. Levels of PAHs in soil and vegetation samples from Tarragona County, Spain. *Environmental Pollution*, 132, 1–11.
- Nadal, M., Schuhmacher, M., Domingo, J. L. 2007. Levels of metals, PCBs, PCNs and PAHs in soils of a highly industrialized chemical/petrochemical area: Temporal trend. *Chemosphere*, 66(2), 267–276. <https://doi.org/10.1016/j.chemosphere.2006.05.020>.
- Nam, J. J., Thomas, G. O., Jaward, F. M., Steinnes, E., Gustafsson, O., Jones, K. C. 2008. PAHs in background soils from Western Europe: Influence of atmospheric deposition and soil organic matter. *Chemosphere*, 70, 1596–1602.
- National Agency of statistics and demography, Senegal (ANDS) 2015. Situation économique et sociale régionale. La région de Kedougou. pp. 13. [French]

- Nesbitt, H.W., Young, G.M., 1982. Early Proterozoic climates and plate motions inferred from major element chemistry of lutites. *Nature* 299, 715–717.
- Niane, B., Moritz, R., Guedron, S., Ngom, PM, Pfeifer, HR, Mall, I., Poté, J. 2014 Effect of recent artisanal small-scale gold mining on the contamination of surface river sediment: case of Gambia River, Kedougou region, south-eastern Senegal. *J Geochem Explor* 144: 517–527.
- Nisbet, I. C. T., LaGoy, P. K. 1992. Toxic equivalency factors (TEFs) for polycyclic aromatic hydrocarbons (PAHs). *Regulatory Toxicology and Pharmacology*, 16, 290–300.
- Nizzetto, L., Cassani, C., Di Guardo, A., 2006. Deposition of PCBs in mountains: the forest filters effect of different forest ecosystem types. *Ecotoxicology and Environmental Safety* 63, 75-83.
- Nziguheba, G., Smolders, E., 2008. Inputs of trace elements in agricultural soils via phosphate fertilizers in European countries. *Sci. Total Environ.* 390, 53–57.
- Olea, R.A., Raju, N.J., Egozcue, J.J., Pawlowsky-Glahn, V., Sing, S., 2018. Advancements in hydrochemistry mapping: Methods and application to groundwater arsenic and iron concentrations in Varanasi, Uttar Pradesh, India. *Stoch. Env. Res. Risk A.* 32 (1), 241–259.
- Oppolzer W. 1991. Intermolecular Diels-Alder Reactions. In: Trost BM, Fleming I (eds) *Comprehensive Organic Synthesis*. Pergamon Press, Oxford, vol 5 chap 4.1.
- Otero, N., Tolosana-Delgado, R., Soler, A., Pawlowsky-Glahn, V., Canals, A., 2005. Relative vs. absolute statistical analysis of compositions: a comparative study of surface waters of a Mediterranean river. *Water Res.* 39, 1404–1414.
- Ottesen, R.T., Birke, M., Finne, T.E., Gosar, M., Locutura, J., Reimann, C., Tarvainen, T., the GEMAS Project Team, 2013. Mercury in European agricultural and grazing land soils. *Appl. Geochem.* 33, 1–12.
- Pacyna, J.M., Breivik, K., Münch, J., Fudala, J., 2003. European atmospheric emissions of selected persistent organic pollutants, 1970-1995. *Atmospheric Environment*. 37, S119-S131.
- Palarea-Albaladejo, J., Martín-Fernandez, J.A., 2013. Values below detection limit in compositional chemical data. *Anal. Chim. Acta* 764, 32–43.
- Pandey, P. K., Patel, K. S., Lenicek, J. 1999. Polycyclic aromatic hydrocarbons: Need for assessment of health risks in India? Study of an urban-industrial location in India. *Environmental Monitoring and Assessment*, 59, 287–319.
- PANNA 2008. Endosulfan Monograph <http://www.panna.org/files/PAN%20Int%20Endosulfan%20Monograph.pdf>>.
- Paone, A., Ayuso, R.A., De Vivo, B., 2001. A metallogenic survey of alkalic rocks of Mt. Somma-Vesuvius volcano. *Mineral. Petrol.* 73, 201–233.

- Parlar, H., Hustert, K., Gab, S., Korte, F., 1979. Isolation, identification, and chromatographic characterization of some chlorinated ClO hydrocarbons in technical chlordane. *J Agric Food Chem* 27, 278-283.
- Parsa, M., Maghsoudi, A., Yousefi, M., Carranza, J.M., 2017. Multifractal interpolation and spectrum-area fractal modelling of stream sediment geochemical data: implications for mapping exploration targets. *J. Afr. Earth Sci.* 128, 5–15.
- Pawlowsky-Glahn, V., Buccianti, A., 2011. *Compositional Data Analysis: Theory and Applications*. John Wiley Sons.
- Pawlowsky-Glahn, V., Egozcue, J.J., 2001. Geometric approach to statistical analysis on the simplex. *Stoch. Env. Res. Risk A.* 15 (5), 384–398.
- Pawlowsky-Glahn, V., Egozcue, J.J., 2011. Exploring compositional data with the CoDA dendrogram. *Austrian J. Stat.* 40 (1–2), 103–113.
- Pawlowsky-Glahn, V., Egozcue, J.J., Lovell, D., 2015a. Tools for compositional data with a total. *Stat. Model.* 15, 175–190.
- Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R., 2015b. *Modelling and Analysis of Compositional Data*. John Wiley Sons, pp. 252.
- Pearson, K., 1897. Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc. Roy. Soc. Lond.* 60, 489–498.
- Peccerillo, A., 2001. Geochemical similarities between the Vesuvius, Phlegraean fields and Stromboli volcanoes: petrogenetic, geodynamic and volcanological implications. *Mineral. Petrol.* 73, 93. <http://dx.doi.org/10.1007/s007100170012>.
- Peccerillo, A., 2005. Plio-quadernary volcanism in Italy. In: *Petrology, Geochemistry, Geodynamics*. Springer-Verlag, Berlin Heidelberg, pp. 365 (ISBN 978-3-540-29,092-6).
- Peirson, D.H., Cawse, P.A., Cambray, R.S., 1974. Chemical uniformity of airborne particulate material, and a maritime effect. *Nature* 251, 675–679.
- Petrik, A., Albanese, S., Lima, A., De Vivo, B., 2018a. The spatial pattern of beryllium and its possible origin using compositional data analysis on a high-density topsoil data set from the Campania Region (Italy). *Appl. Geochem.* <http://dx.doi.org/10.1016/j.apgeochem.2018.02.008>.
- Petrik, A., Albanese, S., Lima, A., Jordan, Gy., De Vivo, B., Rolandi, R., Rezza, C., 2018b. Spatial pattern recognition of arsenic in topsoil using high-density regional data. *Spec. Issue Geochem.: Explor. Environ. Anal.* <https://doi.org/10.1144/geochem2017-060>.
- Petrik, A., Jordan, Gy., Albanese, S., Lima, A., Rolandi, R., De Vivo, B., 2017. Spatial pattern analysis of Ni concentration in topsoils in the Campania Region (Italy). *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.gexplo.2017.09.009>.

- Petrik, A., Thiombane, M., Albanese, S., Lima, A., De Vivo, B., 2018c. Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) through a compositional discrimination analysis: a case study on the Campanian topsoil data. *Geoderma* 331, 87–99.
- Pfaender, F. K., and M. Alexander. 1972. Extensive microbial degradation of DDT in vitro and DDT metabolism by natural communities. *J. Agr. Food Chem.* 20, 842-846.
- Pirrone, N., Cinnirella, S., Feng, X., Finkelman, R. B., Friedli, H. R., Leaner, J., Mason, R., Mukherjee, A. B., Stracher, G. B., Streets, D. G., and Telmer, K.: Global mercury emissions to the atmosphere from anthropogenic and natural sources, *Atmos. Chem. Phys.*, 10, 5951–5964.
- Pison, G., Roosseeuw, P.J., Filzmoser, P., Croux, C., 2003. Robust factor analysis. *J. Multivar. Anal.* 84, 145–172.
- Plant, J.A., Klaver, G., Locutura, J., Salminen, R., Vrana, K., Fordyce, F.M., 1996. Forum of European Geological Surveys (FOREGS) geochemistry task group 1994–1996. In:
- Postma, D., 1985. Concentration of Mn and separation from Fe in sediments—I. Kinetics and stoichiometry of the reaction between birnessite and dissolved Fe(II) at 10 °C. *Geochim. Cosmochim. Acta* 49 (4), 1023–1033. [http://dx.doi.org/10.1016/0016-7037\(85\)90316-3](http://dx.doi.org/10.1016/0016-7037(85)90316-3).
- Pozo, K., Palmeri, M., Palmeri, V., Estellano, V.H., Mulder, M.D., Efstathiou, C.I., Sara, G.L., Romeo, T., Lammel, G., Focardi, S., 2016. Assessing persistent organic pollutants (POPs) in the Sicily Island atmosphere, Mediterranean, using PUF disk passive air samplers. *Environ. Sci. Pollut. R.* 23(20), 796-804.
- Prapamontol, T., Stevenson, D. 1991. Rapid method for the determination of organochlorine pesticides in milk. *J. Chromatogr.* 552, 249-257.
- Qiu, X., Zhu, T., Yao, B., Hu, J., Hu, S., 2005. Contribution of dicofol to the current DDT pollution in China. *Environ. Sci. Technol.* 39, 4385-4390.
- Qu C, Doherty A. L., Xing X., Sun W., Albanese A., Lima A., Qi S. and De Vivo B., 2018. Polyurethane foam-based passive air samplers in monitoring persistent organic pollutants: Theory and application. In: “Environmental Geochemistry - Site Characterization, Data Analysis and Case Histories”, De Vivo B., Belkin H.E Lima A, Eds, Elsevier, Chapt. 20, 521-542
- Qu, C., Albanese, S., Chen, W., Lima, A., Doherty, A.L., Piccolo, A., Arienzo, M., Qi, S., De Vivo, B., 2016. The status of organochlorine pesticide contamination in the soils of the Campanian Plain, southern Italy, and correlations with soil properties and cancer risk. *Environ. Pollut.* 216, 500-511.
- Rahn, K.A., 1976. The chemical composition of the atmospheric aerosol. In: Technical Report of the Graduate School of Oceanography. University of Rhode Island, Kingston, R.I., USA.
- Ravindra, K., Wauters, E., Van Grieken, R. 2008. Variation in particulate PAHs levels and their relation with the transboundary movement of the air masses. *Science of the Total Environment*, 396, 100–110.

- Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., O'Connor, P., GEMAS Team, 2014. Chemistry of Europe's agricultural soils — part A: methodology and interpretation of the GEMAS data set. In: *Geologisches Jahrbuch (Reihe B)*. Schweizerbarth, Hannover, pp. 528.
- Reimann, C., Caritat, P. de, 1998. *Chemical Elements in the Environment. Factsheets for the Geochemist and Environmental Scientist*. Springer-Verlag Berlin, Heidelberg.
- Reimann, C., de Caritat, P., 2000. Intrinsic flaws of element enrichment factors (EFs) in environmental geochemistry. *Environ. Sci. Technol.* 34, 5084–5091.
- Reimann, C., De Caritat, P., 2005. Distinguishing between natural and anthropogenic sources for elements in the environment: regional geochemical surveys versus enrichment factors. *Sci. Total Environ.* 337, 91–107.
- Reimann, C., Filzmoser, P., Garrett, R., 2002. Factor analysis applied to regional geochemical data: problems and possibilities. *Appl. Geochem.* 17 (3), 185–206.
- Reimann, C., Filzmoser, P., Garrett, R.G., Dutter, R., 2008. Statistical data analysis explained. In: *Applied Environmental Statistics With R*. Wiley, Chichester, pp. 362. Chemistry of Europe's agricultural soils — part A: methodology and interpretation of the GEMAS data set. In: Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., O'Connor, P., GEMAS Team (Eds.), *Geologisches Jahrbuch (Reihe B)*, Schweizerbarth: Hannover, pp. 528.
- Reimann, C., Flem, B., Fabian, K., Birke, M., Ladenberger, A., Négrel, P., et al. 2012. Lead and lead isotopes in agricultural soils of Europe: The continental perspective. *Applied Geochemistry*, 27, 532–542.
- Reimann, C., Garrett, R.G., 2005. Geochemical background – concept and reality. *Sci. Total Environ.* 350, 12–27.
- Reimann, C., Garrett, R.G., Filzmoser, P., 2005. Background and threshold – critical comparison of methods of determination. *Sci. Total Environ.* 346, 1–16. GEMAS Team, 2014. In: Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., O'Connor, P. (Eds.), *Chemistry of Europe's Agricultural Soils — Part a: Methodology and Interpretation of the GEMAS Data Set*. *Geologisches Jahrbuch (Reihe B)*, Schweizerbarth: Hannover, pp. 528.
- Rezza, C., Petrik, A., Albanese, S., Lima, A., Minolfi, G., De Vivo, B., 2018. Mo, Sn and W patterns in topsoils of the Campania Region, Italy. In: *Special Issue of the Geochemistry: Exploration, Environment and Analysis*, <http://dx.doi.org/10.1144/geochem2017-061>.
- Roi, R., Sabbiani E (eds). 1993. *Inorganic Mercury. CEC criteria document for occupational exposure limit values*. Joint Research Centre, Commission of the European Communities.
- Rolandi, G., Bellucci, F., Heizler, M.T., Belkin, H.E., De Vivo, B., 2003. Tectonic controls on genesis of ignimbrites from the Campanian Volcanic Zone, Southern Italy. In: De Vivo, B., Scandone, R. (Eds.), *Ignimbrites of the Campania Plain, Italy*. *Mineral. Petrol.* 79. pp. 3–31.

- Roosseeuw, P.J., Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41, 212–223.
- Rossini, O.S., Fernández, E.A.J., 2007. Monitoring of heavy metals in topsoil, atmospheric particles and plant leaves to identify possible contamination sources. *Microchem. J.* 86, 131–139.
- Saeedi, M., Li, L.Y., Salmanzadeh, M., 2012. Heavy metals and polycyclic aromatic hydrocarbons pollution and ecological risk assessment in street dust of Tehran. *J. Hazard. Mater.* 227–228, 9–17.
- Salminen, R., Batista, M.J., Bidovec, et al., 2005. FOREGS Geochemical Atlas of Europe, Part 1: Background Information, Methodology and Maps. Geological Survey of Finland, Espoo, pp. 526. <http://weppi.gtk.fi/publ/foregsatlas/>.
- Salminen, R., Gregorauskiene, V., 2000. Considerations regarding the definition of a geochemical baseline of elements in the surficial materials in areas differing in basic geology. *Appl. Geochem.* 15, 647–653.
- Salminen, R., Tarvainen, T., Demetriades, A., Duris, M., Fordyce, F.M., Gregorauskiene, V., et al., 1998. FOREGS Geochemical Mapping Field Manual. Espoo: Guide 47. Geological Survey of Finland.
- Scheib, A.J., Birke, M., Dinelli, E., The Gemas Project Team, 2014. Geochemical evidence of Aeolian deposits in European soils. *Boreas* 43 (1), 175–192. <http://dx.doi.org/10.1111/bor.12029>.
- Senthil Kumar, K., Kannan, K., Subramanian, A.N., Tanabe, S., 2001. Accumulation of organochlorine pesticides and polychlorinated biphenyls in sediments, aquatic organisms, birds and bird eggs and bat collected from South India. *Environmental Science and Pollution Research.* 8, 35–47.
- Silva, M.H., Carr Jr., W.C., 2009. Human health risk assessment of endosulfan. II: Dietary exposure assessment. *Regul. Toxicol. Pharmacol.* 56, 18–27.
- Simcik, M. F., Eisenreich, S. J., Lioy, P. J. 1999. Source apportionment and source/sink relationships of PAHs in the coastal atmosphere of Chicago and Lake Michigan. *Atmospheric Environment*, 33(30), 5071–5079. [https://doi.org/10.1016/S1352-2310\(99\)00233-2](https://doi.org/10.1016/S1352-2310(99)00233-2).
- Sinclair, A.J., 1974. Selection of threshold values in geochemical data using probability graphs. *J. Geochem. Explor.* 3, 129–149.
- Sinclair, A.J., 1976. Application of probability graphs in mineral exploration. In: Special Volume 4. Association of Exploration Geochemists, Toronto, pp. 95.
- Sinclair, A.J., 1983. Univariate analysis. Chapter 3. In: Howarth, R.J. (Ed.), *Statistics and Data Analysis in Geochemical Prospecting*. In: G.J.S. (Series Editor), *Handbook of Exploration Geochemistry*, vol. 2. Elsevier, Amsterdam, pp. 59–81.
- Skordas, K., Kelepertsis, A., 2005. Soil contamination by toxic metals in the cultivated region of Agia, Thessaly, Greece. Identification of sources of contamination. *Environ. Geol.* 48 (4), 615–624. <https://doi.org/10.1007/s00254-005-1319-x>.

- Soclo, H. H., Garrigues, P., Ewald, M. 2000. Origin of polycyclic aromatic hydrocarbons (PAHs) in coastal marine sediments: Case studies in Cotonou (Benin) and Aquitaine (France) areas. *Marine Pollution Bulletin*, 40(5), 387–396. [https://doi.org/10.1016/S0025-326X\(99\)00200-3](https://doi.org/10.1016/S0025-326X(99)00200-3).
- Sofowote, U. M., McCarry, B. E., Marvin, C. H. 2008. Source apportionment of PAH in Hamilton Harbour suspended sediments: Comparison of two factor analysis methods. *Environmental Science and Technology*, 42(60), 07–14.
- Spaziani, F., Angelone, M., Coletta, A., Salluzzo, A., Cremisini, C. (2008). Determination of platinum group elements and evaluation of their traffic-related distribution in Italian urban environments. *Analytical Letters*, 41(14), 2658–2683. <https://doi.org/10.1080/00032710802363503>.
- SSR. 2010. Soil Contamination in West Africa | Environmental Remediation Pollution. (also available at <https://www.scribd.com/doc/71599035/Soil-Contamination-in-West-Africa>).
- Stockholm Convention on Persistent Organic Pollutants (POPs), (2011). <http://chm.pops.int/>.
- Stockholm Convention on Persistent Organic Pollutants (POPs). (2005). World Wide Found, Stockholm Convention “New POPs”: screening additional POPs candidates, April 2005, pp. 38.
- Suchan, P., Pulkrabová, J., Hajslová, J., Kocourek, V., (2004). Pressurized liquid extraction in determination of polychlorinated biphenyls and organochlorine pesticides in fish samples. *Anal. Chim. Acta*. 520,193–200.
- Sucharová, J., Suchara, I., Hola, M., Marikova, S., Reimann, C., Boyd, R., Filzmoser, P., Englmaier, P., 2012. Top-/Bottom-soil ratios and enrichment factors: what do they really show? *Appl. Geochem*. 27, 138–145.
- Sutherland, R.A., Tolosa, C.A., Tack, F.M.G., Verloo, M.G., 2000. Characterization of selected element concentration and enrichment ratios in background and anthropogenically impacted roadside areas. *Arch. Environ. Contam. Toxicol*. 38, 428–438.
- Sylla, M., Ngom, P.M., 1997. Le gisement d'or de Sabodala (Sénégal Oriental): Une minéralisation filonienne d'origine hydrothermale remobilisée par une tectonique cisailante. *J. Afr. Earth Sci*. 25, 183–192. [French]
- Szeto, S.Y., Price, P.M., (1991). Persistence of pesticide residues in mineral and organic soils in the Fraser valley of British Columbia. *J. Agric. Food Chem*. 39, 1679–1684.
- Tang, L., Tang, X., Zhu, Y., Zheng, M., Miao, Q. (2005). Contamination of polycyclic aromatic hydrocarbons (PAHs) in urban soils in Beijing, China. *Environment International*, 31(6), 822–828. <https://doi.org/10.1016/j.envint.2005.05.031>.
- Tariq, R.S., Shah, M.H., Shaheen, N., Khalique, K., Manzoor, M., Jaffar, S., 2005. Multivariate analysis of trace metal levels in tannery effluents in relation to soil and water: a case study from Peshawar, Pakistan. *J. Environ. Manag*. 79, 20–29.

- Tarvainen, T., Jarva, J., Johnson, C.C., Ottesen, R.T., 2011. Using geochemical baselines in the assessment of soil contamination in Finland. In: Demetriades, A., Locutura, J. (Eds.), *Mapping the Chemical Environment of Urban Areas*. Chichester, UK, John Wiley Sons Ltd., pp. 223–231.
- Tarzia, M., De Vivo, B., Somma, R., Ayuso, R.A., McGill, R.A.R., Parrish, R.R., 2002. Anthropogenic versus natural pollution: an environmental study of an industrial site under remediation (Naples, Italy). *Geochem. Explor. Environ. Anal.* 2, 45–56.
- Taylor, S.R., McLennan, S.M., 1995. The geochemical evolution of the continental crust. *Rev. Geophys.* 33, 241–265.
- Telmer K, Veiga M. (2008). World emissions of mercury from artisanal and small scale gold mining. In: Pirrone N, Mason R, editors. Interim Report of the UNEP Global Partnership on Atmospheric Mercury Transport and Fate Research. UNEP; 2008. [<http://www.mercurywatch.org/userfiles/file/Telmer%20and%20Veiga%202009%20Springer.pdf>].
- Templ, M., Hron, K., Filzmoser, P., 2011. *Rob-Compositions: Robust Estimation for Compositional Data. Manual and Package, Version 1.4.4*.
- Tennant, C.B., White, M.L., 1959. Study of the distribution of some geochemical data. *Econ. Geol.* 54, 1281–1290.
- Thiombane, M., Martin-Fernandez, J.A., Albanese, S., Lima, A., Doherty, A., De Vivo, B., 2018b. Exploratory analysis of multi-element geochemical patterns in soil from the Sarno River Basin (Campania region, southern Italy) through compositional data analysis (CODA). *J. Geochem. Explor.* 195, 110–120.
- Thiombane, M., Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Cavaliere, M., De Vivo, B., 2018a. Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model. *J. Geochem. Explor.* 189, 85–99.
- Tobiszewski, M., Namieśnik, J. (2012). PAH diagnostic ratios for the identification of pollution emission sources. *Environmental Pollution*, 162, 110–119.
- Tolosana-Delgado, R., van den Boogaart, K.G., 2013. Joint consistent mapping of high dimensional geochemical surveys. *Math. Geosci.* 45, 983–1004.
- Torrente, M., Milia, A., 2013. Volcanism and faulting of the Campania margin (Eastern Tyrrhenian Sea, Italy): a three-dimensional visualization of a new volcanic field off Campi Flegrei. *Bull. Volcanol.* 75, 719. <https://doi.org/10.1007/s00445-013-0719-0>.
- Tóth, G., Guicharnaud, R.-A., Tóth, B., Hermann, T., 2014. Phosphorus levels in croplands of the European Union with implications for P fertilizer use. *Eur. J. Agron.* 55, 42–52. (ISSN 1161-0301). <https://doi.org/10.1016/j.eja.2013.12.008>.
- Tukey, J.W., 1977. *Exploratory Data Analysis*. Addison-Wesley Publishing Company. Varrica, D., Tamburo, E., Milia, N., Vallascas, E., Cortimiglia, V., De Giudici, G., Dongarrà, G., Sanna, E., Monna, F., Losno, R., 2014. Metals and metalloids in hair samples of children living near the

- abandoned mine sites of Sulcis-Iglesiente (Sardinia, Italy). *Environ. Res.* 134 (Suppl. C), 366–374. <https://doi.org/10.1016/j.envres.2014.08.013>.
- U.S. Environmental Protection Agency (EPA), 2001. Guidance for characterizing background chemicals in soil at superfund sites, external review draft, Office of Emergency and Remedial Response, OSWER. 9285.7-41. In: Replaced by Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites, EPA 540-R-01-003, September 2002.
- UNECE (United Nation Economic Commission for Europe). 2010. Convention on long-range transboundary air pollution. http://www.unece.org/env/lrtap/pops_h1.htm.
- United Nations Environment Programme (UNEP) 2013. Minamata Convention agreed by nations. <http://www.unep.org/newscentre/default.aspx?DocumentID=2702&ArticleID=9373>.
- USEPA 2002. US Environmental Protection Agency Guidance on environmental data verification and data validation (p. 96). EPA QA/G-8. Washington, DC: Office of Environmental Information.
- USEPA, 1991. Risk Assessment Guidance for Superfund: Volume I — Human Health Evaluation Manual (Part B, Development of Risk-Based Preliminary Remediation Goals). U.S. Environmental Protection Agency, Washington, DC. [EPA/540/R-92/003]
- USEPA, 2000. Ultrasonic Extraction, Test Methods for Evaluating Solid Waste, Method 3550C, Revision 3, US Environmental Protection Agency, Washington, DC.
- USEPA, 2007. Endosulfan. Hazard Characterization and Endpoint Selection Reflecting Receipt of a Developmental Neurotoxicity Study and Subchronic Neurotoxicity Study. PC Code: 079401. DP Barcode D338576. Reaves, E., Washington, DC: Office of Prevention, Pesticides and Toxic Substances, U.S. Environmental Protection Agency. April 7, 2007. <http://www.epa.gov/pesticides/reregistration/endosulfan/>.
- Van Den Boogaart, K.G., Tolosana-Delgado, R., Bren, R., 2011. Compositions: Compositional Data Analysis. R Package Version 1. pp. 10–12. Available at: <http://CRAN.R-project.org/package=compositions>.
- Vane, C. H., Kim, A. W., Beriro, D. J., Cave, M. R., Knights, K., Moss-Hayes, V., et al. 2014. Polycyclic aromatic hydrocarbons (PAH) and polychlorinated biphenyls (PCB) in urban soils of Greater London, UK. *Applied Geochemistry*, 51, 303–314.
- Varrica, D., Tamburo, E., Milia, N., Vallascas, E., Cortimiglia, V., De Giudici, G., Dongarrà, G., Sanna, E., Monna, F., Losno, R., 2014. Metals and metalloids in hair samples of children living near the abandoned mine sites of Sulcis-Iglesiente (Sardinia, Italy). *Environmental Research*, 134 (Supplement C), 366–374. doi:<https://doi.org/10.1016/j.envres.2014.08.013>.
- Veiga, M.M., Baker, R., 2004. Protocols for environmental and health assessment of mercury released by artisanal and small-scale gold miners. Global Mercury Project. Vienna, pp. 289.
- Vercoutere, K., Fortunati, U., Muntau, H., Griepink, B., Maier, E.A. 1995. The certified reference materials CRM 142 R light sandy soil, CRM 143 R sewage sludge amended soil and CRM145R

- sewage sludge for quality control in monitoring environmental and soil pollution. *Fres. J. Anal. Chem.* 352:197-202.
- Villanneau, E.J., Saby, N.P.A., Marchant, B.P., Jolivet, C., Boulonne, L., Caria, G., Barriuso, E., Bispo, A., Briand, O., Arrouays, D., 2011. Which persistent organic pollutants can we map in soil using a large spacing systematic soil monitoring design? A case study in Northern France. *Science of the Total Environment* 409, 3719-3731.
- Vitale, S., Ciarcia, S., 2013. Tectono-stratigraphic and kinematic evolution of the southern Apennines/Calabria-Peloritani Terrane system (Italy). *Tectonophysics* 583, 164–182.
- Vitale, S., Ciarcia, S., Mazzoli, S., Iannace, A., Torre, M., 2010. Structural analysis of the Internal Units of Cilento, Italy: new constraints of the Miocene tectonic evolution of the southern Apennine accretionary wedge. *Compt. Rendus Geosci.* 342, 475–482.
- Walker, K., Vallero, D.A., Lewis, R.G., 1999. Factors influencing the distribution of lindane and other hexachlorocyclohexane in the environment. *Environ. Sci. Technol.* 33, 4373–4378.
- Wang, Y., Sikora, S., Kim, H., Dubey, B., Townsend, T., 2012. Mobilization of iron and arsenic from soil by construction and demolition debris landfill leachate. *Waste Manag.* 32 (5), 925–932. <https://doi.org/10.1016/j.wasman.2011.11.016>.
- Washington, H.S., 1906. The Roman Comagmatic Region, vol. 57. Carnegie Institution of Washington, pp. 199.
- Wedepohl, K.H. (Ed.), 1978. *Handbook of Geochemistry*. Springer-Verlag, Berlin-Heidelberg.
- Wedepohl, K.H., 1995. The composition of the continental crust. *Geochim. Cosmochim. Acta* 59, 1217–1232.
- Weinberg, J. 1998. Overview of POPs and need for a POPs treaty. Public forum on persistent organic pollutants-the international POPs elimination network. 1998.
- WHO, 2003. Environmental Mercury and Inorganic Mercury Compounds: Human health Aspects, Accessed April 2015 <http://www.who.int/ipcs/publications/cicad/en/cicad50.pdf>.
- Winship KA. (1986). Organic mercury compounds and their toxicity. *Adv Drug React Ac Pois Rev* 1986;3: 141–180.
- Wu, S., Peng, S., Zhang, X., Wu, D., Luo, W., Zhang, T., Zhou, S., Yang, G., Wan, H., Wu, L., 2015. Levels and health risk assessments of heavy metals in urban soils in Dongguan, China. *J. Geochem. Explor.* 148, 71–78. <https://doi.org/10.1016/j.gexplo.2014.08.009>.
- Wu, S., Zhou, S., Li, X., 2011. Determining the anthropogenic contribution of heavy metal accumulations around a typical industrial town: Xushe, China. *J. Geochem. Explor.* 110, 92–97.
- Yadav, I. C., Devi, N. L., Li, J., Zhang, G., and Breivik, K.: Possible emissions of POPs in plain and hilly areas of Nepal: Implications for source apportionment and health

- risk assessment, *Environ. Pollut.*, 220, 1289–1300.
<https://doi.org/10.1016/j.envpol.2016.10.102>, 2017.
- Yang, D., Qi, S.H., Zhang, J.Q., Tan, L.Z., Zhang, J.P., Zhang, Y., Xu, F., Xing, X.L., Hu, Y., Chen, W., Yang, J.H., Xu, M.H., 2012. Residues of organochlorine pesticides (OCPs) in agricultural soils of Zhangzhou City, China. *Pedosphere*. 22, 178–189.
- Yang, J.Y., Yang, X.E., He, Z.L., Chen, G.C., Shentu, J.L., Li, T.Q., 2004. Adsorption–desorption characteristics of lead in variable charge soils. *J. Environ. Sci. and Health, Part A* 39 (8), 1949–1967.
- Yang, X. L., Wang, S. S., Bian, Y. R., Chen, F., Yu, G. F., Gu, C. G., 2008. Dicofol application resulted in high DDTs residue in cotton fields from Northern Jiangsu province, China. *Journal of hazardous materials*. 150, 92–98.
- Yunker, M. B., Macdonald, R. W., Vingarzan, R., Mitchell, R. H., Goyette, D., Sylvestre, S. 2002. PAHs in the Fraser river basin: A critical appraisal of PAH ratios as indicators of PAH source and composition. *Organic Geochemistry*, 33, 489–515.
- Z.L. Zhang, J. Huang, G. Yu, et al., 2004. Occurrence of PAHs, PCBs and organochlorine pesticides in Tonghui River of Beijing, China, *Environ. Pollut.* 130, 249–261.
- Zhang, H. B., Luo, Y. M., Wong, M. H., Zhao, Q. G., Zhang, G. L. 2006. Distributions and concentrations of PAHs in Hong Kong soils. *Environmental Pollution*, 141, 107–114.
- Zhao, L., Xu, Y., Hou, H., Shangguan, Y., Li, F., 2014. Source identification and health risk assessment of metals in urban soils around the Tanggu chemical industrial district, Tianjin, China. *Sci. Total Environ.* 468 (Suppl. C), 654–662. <https://doi.org/10.1016/j.scitotenv.2013.08.094>.
- Zhao, S., Feng, C., Wang, D., Liu, Y., Shen, Z., 2013. Salinity increases the mobility of Cd, Cu, Mn, and Pb in the sediments of Yangtze Estuary: relative role of sediments' properties and metal speciation. *Chemosphere* 91 (7), 977–984. <https://doi.org/10.1016/j.chemosphere.2013.02.001>.
- Zhou, Q., Wang, J., Meng, B., Cheng, J., Lin, G., Chen, J., Zheng, D., Yu, Y., 2013. Distribution and sources of organochlorine pesticides in agricultural soils from central China. *Ecotoxicol. Environ. Saf.* 93, 163–170.
- Zuo, R., 2011. Decomposing of mixed pattern of arsenic using fractal model in Gangdese belt, Tibet, China. *Appl. Geochem.* 26, S271–S273.
- Zuo, R., Wang, J., 2016. Fractal/multifractal modelling of geochemical data: a review. *J. Geochem. Explor.* 164, 33–41. Zuo, R., Wang, J., Chen, G., Yang, M., 2015. Identification of weak anomalies: a multifractal perspective. *J. Geochem. Explor.* 148, 12–24.
- Zuo, R., Wang, J., Chen, G., Yang, M., 2015. Identification of weak anomalies: a multifractal perspective. *J. Geochem. Explor.* 148, 12–24.
- Zuo, R., Xia, Q., Wang, H., 2013. Compositional data analysis in the study of integrated geochemical anomalies associated with mineralization. *Appl. Geochem.* 28, 202–211.

- Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Zuo, R., De Vivo, B., 2018. Exploring uni-element geochemical data under a compositional perspective. *Appl. Geochem.* <http://dx.doi.org/10.1016/j.apgeochem.2017.10.003>.
- Zuzolo, D., Cicchella, D., Catani, V., Giaccio, L., Guagliardi, I., Esposito, L., De Vivo, B., 2017. Assessment of potentially harmful elements pollution in the Calore River basin (southern Italy). *Environ. Geochem. Health* 39, 531–548. <https://doi.org/10.1007/s10653-016-9832-2>.

Appended publications

Paper 1

Thiombane, M., Di Bonito, M., Albanese, A., Zuzolo, D., Lima, A., De Vivo, D. (2018). Geogenic versus anthropogenic behaviour of geochemical phosphorus footprint in the Campania region (Southern Italy) soils through compositional data analysis and enrichment factor. *Geoderma*. Volume 335, 1 February 2019, Pages 12-26. <https://doi.org/10.1016/j.geoderma.2018.08.008>.

Paper 2

Petrik, A., **Thiombane, M.,** Lima, A., Albanese, S., De Vivo. B. (2018). Soil Compositional Contamination Index: a new approach to reveal and quantify contamination patterns of potentially toxic elements in Campania Region (Italy), *Journal of applied geochemistry*. 96, 264-276. <https://doi.org/10.1016/j.apgeochem.2018.07.014>.

Paper 3

Thiombane, M., Petrik, A., Albanese, S., Lima, A., De Vivo. B. (2018). Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) through a compositional discrimination analysis: a case study on the Campanian topsoil data. *Geoderma*, 331, 87-99. <https://doi.org/10.1016/j.geoderma.2018.06.019>.

Paper 4

Thiombane M., Albanese, S., Martín-Fernández, J., Lima A., Doherty A. De Vivo, B., (2018). Exploratory analysis of multi-element geochemical patterns in soil from the Sarno River Basin (Campania region, southern Italy) through compositional data analysis (CODA). *Journal of Geochemical exploitation*, 195, 110-120. <https://doi.org/10.1016/j.gexplo.2017.06.010>.

Paper 5

Thiombane, M., Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Cavaliere, M., De Vivo.B., (2018). Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model. *Journal of Geo Exploration*. 172, 174-183. <https://doi.org/10.1016/j.gexplo.2018.03.010>.

Paper 6

Thiombane, M., Albanese, S, Di Bonito, M., Lima, A., Rolandi, R., Qi, S., De Vivo. B. (2018). Source patterns and characterisation of Polycyclic Aromatic hydrocarbons (PAHs) in urban and rural areas of Southern Italy. *Journal of environmental geochemistry and health*. <https://doi.org/10.1007/s10653-018-0147-3>.

Paper 7

Thiombane, M., Petrik, A., Di Bonito, M., Albanese, S., Zuzolo, D., Cicchella, D., Lima, A., Qu, C., Qi, S., De Vivo B., (2018). Status, sources and contamination levels of organochlorine pesticides residues in urban and agricultural areas: A preliminary review in central-southern Italian soils. *Journal of Environmental Science and Pollution Research*. <https://doi.org/10.1007/s11356-018-2688-5>.

Paper 1

Geogenic versus anthropogenic behaviour of geochemical phosphorus footprint in the Campania region (Southern Italy) soils through compositional data analysis and enrichment factor

Matar Thiombane, Marcello Di Bonito, Stefano Albanese, Daniela Zuzolo, Annamaria Lima, Benedetto De Vivo

Journal of Geoderma, Volume 335, 1st February 2019, Pages 12-26



Geogenic versus anthropogenic behaviour and geochemical footprint of Al, Na, K and P in the Campania region (Southern Italy) soils through compositional data analysis and enrichment factor

Matar Thiombane^{a,*}, Marcello Di Bonito^b, Stefano Albanese^a, Daniela Zuzolo^c, Annamaria Lima^a, Benedetto De Vivo^{d,e}

^a Department of Earth, Environment and Resources Sciences (DiSTAR), University of Napoli "Federico II", Complesso Universitario di Monte Sant'Angelo, Via Cintia snc, 80125 Napoli, Italy

^b School of Animal, Rural and Environmental Sciences, Brackenhurst Campus Southwell NG25 0QF Nottingham Trent University, United Kingdom

^c Department of Science and Technology, University of Sannio, via dei Mulini 59/A, 82100 Benevento, Italy

^d Pegaso University, Piazza Trieste e Trento 48, 80132 Naples, Italy

^e Benecon Scarl, Dip. Ambiente e Territorio, Via S. Maria di Costantinopoli 104, 80138 Naples, Italy

ARTICLE INFO

Handling Editor: Edward A Nater

Keywords:

Compositional data
Multi-fractal computation
Robust biplot
Background concentration
Enrichment factor

ABSTRACT

Geochemical studies that focus on environmental applications tend to approach the chemical elements as individual entities and may therefore offer only partial and sometimes biased interpretations of their distributions and behaviour. A potential alternative approach is to consider a compositional data analysis, where every element is part of a whole. In this study, an integrated methodology, which included compositional data analysis, multifractal data transformations and interpolation, as well as enrichment factor analysis, was applied to a geochemical dataset for the Campania region, in the south of Italy, focusing in particular on the behaviour, footprints and sources of a smaller pool of elements: Al, Na, K and P. The initial dataset included 3669 topsoil samples, collected at an average sampling density of 1 site per 2.3 km², and analyzed (after an aqua regia extraction) by a combination of ICP-AES and ICP-MS for 53 elements. Frequency based methods (Clr biplot, Enrichment Factor computation) and frequency spatial-method (fractal and multifractal plots) allowed identifying the relationships between the elements and their possible source patterns in Campania soils in relation to a natural occurring concentrations in geogenic material (rocks, soils and sediments) or human input. Results showed how the interpretation of concentration and behaviour of Al, Na, K and P was enhanced thanks to the application of data log-ratio transformation in univariate and multivariate analysis compared to the use of raw or log-normal data. Multivariate analyses with compositional biplot allowed the identification of four element associations and their potential association with the underlying geology and/or human activities. When focusing on the smaller pool of elements (Al, P, K and Na), these relationships with the unique geology of the region, were largely confirmed by multifractal interpolated maps. However, when the local background was used for the calculation of the enrichment factor, the resulting interpolated maps allowed to identify smaller areas where the greater concentrations of P could not be possibly associated to a mineralisation (e.g., ultrapotassic rocks) but were more likely to be associated to anthropogenic input such as agriculture activities with potentially extensive use of phosphate fertilizers. The integrated approach of this study allowed a more robust qualitative and quantitative evaluation of elemental concentration, providing in particular new and vital information on the distribution and patterns of P in soils of the Campania region, but also a viable, more robust, methodological approach to regional environmental geochemistry studies.

1. Introduction

Element distributions in soils are generally related to a variety of factors such as geology, chemical reactivity, mineralogy, hydrology,

vegetation, and anthropogenic activities. In recent years environmental geochemistry has allowed to gain much insight on the relationship between chemical elements in soils and their sources by means of various tools. In particular, univariate and multivariate analysis (Otero

* Corresponding author.

E-mail address: thiombane.matar@unina.it (M. Thiombane).

<https://doi.org/10.1016/j.geoderma.2018.08.008>

Received 31 October 2017; Received in revised form 28 June 2018; Accepted 6 August 2018

0016-7061/ Crown Copyright © 2018 Published by Elsevier B.V. All rights reserved.

et al., 2005; Reimann et al., 2008; Zuo, 2011; Thiombane et al., 2018a) as well as spatial analysis (frequency space-method) have helped to discriminate between anthropogenic contribution (and contamination) compared to natural or geogenic sources. More recently, different approaches have also been used to treat compositional geochemical data in a more comprehensive way. Geochemical data are closed number systems (parts of a whole) and they should be treated as that to avoid spurious correlations and misleading interpretations. By taking only the absolute values or the log-normal data, the proportional nature of the geometry of the simplex may be not fully captured (Han and Kamber, 2001). Compositional data instead, consider each element as part of a whole which carry relative information (Aitchison, 1986; Egozcue et al., 2003). For this reason, and to minimize and/or eliminate the presence of outliers and spurious correlation (Pawlowsky-Glahn and Buccianti, 2011), they have increasingly been approached in alternative ways. In particular, the use of log-ratio transformations such as additive log-ratio (alr), centered log-ratio (clr) (Aitchison, 1986) and isometric log-ratio (ilr) (Egozcue et al., 2003) were found to be effective approaches to deal with these complex datasets and better explain their significance. Furthermore, mapping remains a useful tool to display data distribution through fractal and multifractal analysis (Mandelbrot, 1983; Cheng et al., 1994) and through geostatistics (Olea et al., 2018; Thiombane et al., 2018b). These have been demonstrated to be a powerful means for identifying geochemical anomalies (Cheng, 1999, 2007; Cheng et al., 2000, 2010; Lima et al., 2003a; Cicchella et al., 2005) and enhance elements patterns.

Aside from their inherent complexity, often environmental geochemistry studies still present some discrepancies in the definition of key concepts and therefore interpretation of results. For example, the concept of background and/or baseline concentration are still confused and misused, whilst they should be clearly defined and interpreted: a background value corresponds to the range of concentration of a given element in a given area which is completely dependent on the compositional and mineralogical characteristic of the parent/source geological material (Reimann et al., 2005; Albanese et al., 2007). On the other hand, a baseline value relates the actual most diffused range of concentration of a given element in a specific area depending both on the nature of the parent geological/source material (Salminen and Gregorauskiene, 2000) and on the historic diffuse release into the environment from anthropogenic sources.

In reality, in areas where the anthropogenic impact is very small or not significant, the background and baseline values can overlap, as it is often difficult to distinguish the natural sources compared to the anthropogenic contribution. In areas where the anthropogenic impact is evident or significant, however, the background and baseline values should be interpreted appropriately and in a distinct way. This study focuses on a region in south of Italy, Campania, characterized by complex geological and geomorphological features and by a significant presence of human activities, such as industry, agriculture, tourism and urbanization. The industrial presence is often related to the agricultural activities and is mostly developed in proximity of the agricultural and surrounding urban areas. Many of these industrial activities are devoted to food production and preservation process (e.g., San Marzano tomatoes conserves), but also to clothes productions and tanneries (e.g., Solofra tanneries industries). These activities represent the main livelihood for the local communities and a major economic input not only for the Campania region, but for the Italian economy. Nevertheless, as with any agriculture activity, it is inevitable that chemical fertilizers (phosphates and sulfates based) are used to improve soil productivity and quality. Some of the benefits of using fertilizers, however, carry also 'externalities' that can travel from the soil, to the crop, and the entire food chain. In particular, metallic impurities of P-based fertilizers containing Cu, As and Cd, can affect the quality of the soil and possibly its contamination; phosphate fertilizers are in fact known as a major source of trace metals among all mineral fertilizers (Nziguheba and Smolders, 2008). These sources of contamination can be investigated by

means of Enrichment Factor (Chester and Stoner, 1973) which is able to display the degree of contamination related to a mineralisation (geogenic) or anthropogenic activities (Reimann and de Caritat, 2005).

This study proposes an assessment of the potential impacts and footprint of some of the agricultural activities in Campania, by use of compositional data analysis, concentration-area (C-A) and spectrum-areas (S-A) fractal and multifractals models applied to geochemical data. In particular, the study focuses on a smaller pool of elements that can be potentially directly related to either geogenic, anthropogenic or mixed source: Al, Na, K, and P will be investigated to detect their potential origin and their background concentrations in Campania soils. The main objectives of this study were:

(1) to illustrate the importance of compositional log-transformations on geochemical data; (2) to delineate the main sources of elements by using a combination of multivariate analysis and GIS based approach; (3) to use enrichment factor (EF) to assess geogenic or anthropogenic behaviour of P in the study area; (4) to prove the importance and ease of using the local reference elements as opposed to the traditional continental crust references values in EF calculation to determine degree of contamination. The results from this investigation could greatly influence and provide a blueprint to future similar studies.

2. Materials and methods

2.1. Features of the study areas

The Campania region is located in the south of Italy, covering an area of about 13,600 km². The region borders the Tyrrhenian Sea and the Lazio region at the western and northern sides, respectively (Fig. 1-A).

The main geological features of the Campania region are constituted by the Apennines chain which cross the areas as a backbone oriented NW-SE. The hilliest part of the chain forms the Mt. Matese in the north, Mt. Taburno and Picentini in the center, and the Mt. Alburni in the southeast. These mountains are mostly formed by sedimentary rocks such as limestone and dolostone whilst the external domains are constituted by siliceous schist and terrigenous sediments (clays, siltstone, sandstone, and conglomerate) (Bonardi et al., 1998; De Vivo et al., 2016). Igneous rocks are present in the region and are mostly formed by potassic and ultrapotassic volcanic rocks of the Roccamonfina volcano (De Vivo et al., 2001, 2010, 2016; Rolandi et al., 2003; Albanese et al., 2007) in the northwest, the Mt. Somma-Vesuvius, Campi Flegrei and Ischia volcanoes which are located nearer the coastal areas of the study region. The coastal areas and plains are constituted by alluvial, lacustrine and coastal (mixes of oceanic and terrestrial) sediments (Bonardi et al., 1998; De Vivo et al., 2016).

The hydrography of the Campania region is characterized by three main river catchments (Garigliano, Volturno and Sele) and by numerous minor streams, mostly draining towards the Tyrrhenian Sea. The morphology of the drainage network is irregular and controlled by geological and structural features and rivers are source of water irrigation for agriculture field (Ducci and Tranfaglia, 2005).

Campania is one of the most populated regions of Italy with > 5.8 million inhabitants (ISTAT, 2016). This high density of population is coupled by the presence of a large number of industrial activities, with the majority being involved in agriculture: vineyards and olive plantations – mostly in hilly areas – seasonal crops, and greenhouse products (tomatoes, potatoes, aubergines, peppers, peas, and citrus fruits) represent major resources for the region and the local economy (Albanese et al., 2007). This intensive agriculture activity occupies > 50% of the total land and occurs mostly in the coastal and mountainous areas (Fig. 1B), where fertile land and suitable soils are occurring. Unfortunately, such industrial activities are known to have a potential negative impact if not properly managed, contributing to the contamination of natural resources such as superficial and groundwater as well as soils. Campania is not immune to these problems, and some

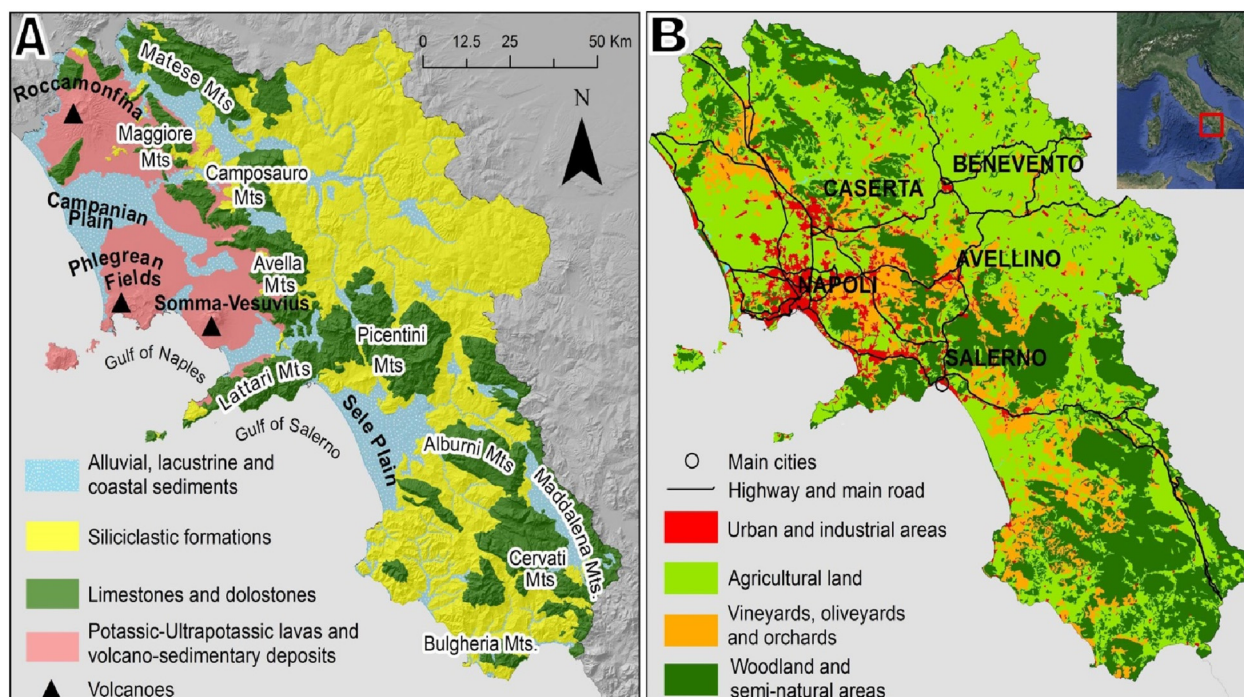


Fig. 1. Simplified geological map (A) and land use (B) of the Campania region (southern of Italy).

studies have already highlighted their existence and relation to its natural resources (Cicchella et al., 2005; Minolfi et al., 2018).

2.2. Sampling procedures and analyses

From 2013 to 2015, 3669 samples were collected from topsoil of the Campania Region (13,600 km²) with a nominal density of 1 sample in each 3.2 km². Each top soil sample (from 0 to 20 cm) was made by homogenizing 5 subsamples at the corners and the centre of a 100m² square, collecting approximately 1.5 kg in total. The sampling procedure followed the Geochemical Mapping of Agricultural and Grazing Land Soils (GEMAS) sampling procedure described by Reimann et al. (2014). At each sampling site, several physico-chemical parameters of the soil properties were measured, including pH, total water content, conductivity, total organic content and the geographical coordinates system recorded by geospatial positioning systems (GPS).

Chemical analyses were carried out at an international accredited Laboratory, Acme Analytical Laboratories Ltd. (now Bureau Veritas, Vancouver, Canada). The samples were analyzed after an aqua regia extraction, by a combination of inductively coupled plasma atomic emission (ICP-AES) and inductively coupled plasma mass spectrometry (ICP/MS) for “pseudototal” concentration of 53 elements (Ag, Al, As, Au, B, Ba, Be, Bi, Ca, Cd, Ce, Co, Cr, Cs, Cu, Fe, Ga, Ge, Hf, Hg, In, K, La, Li, Mg, Mn, Mo, Na, Nb, Ni, P, Pb, Pd, Pt, Rb, Re, S, Sb, Sc, Se, Sn, Sr, Ta, Te, Th, Ti, Tl, U, V, W, Y, Zn, and Zr). A sub-sample of 15 g of the sieved < 2 mm soil fraction was digested in 90 ml aqua regia and leached for 1 h in a 95 °C water bath. After cooling, the solution was diluted to a final volume of 300 ml using a solution of 5% HCl. The sample weight to solution volume ratio was 1 g per 20 ml. The solutions were analyzed using a Perkin Elmer Elan 6000/9000 inductively coupled plasma emission mass spectrometer (ICP-MS). The accuracy and precision of the data was measured by comparison to known analytical standards. Calibration solutions were included at the beginning and end of each analytical run (a total of 40 solutions). Precision is ± 100% at the detection limit, and improves to better than ± 10% at concentrations 50 times the detection limit or greater.

2.3. Compositional data analysis

Nowadays it appears necessary to reconsider geochemical data under a compositional data analysis perspective (Aitchison, 1986; Buccianti et al., 2006, 2014, 2018; Pawłowsky-Glahn and Buccianti, 2011). A composition is defined as a sample space of the regular unit D-simplex, S^D that is a vector of D positive components summing up to a given constant k , set typically equal to 1 (proportions), 100 (percentages), or 10⁶ (ppm) by closure. It relates parts of some whole that carry relative information (ratios of components) whose sample space is the simplex (Pawłowsky-Glahn and Egozcue, 2001).

$$S^D = \left\{ X = [x_1, x_2, \dots, x_D] \mid x_i > 0; \sum_{i=1}^D x_i = k \right\} \quad (1)$$

As explained in the introduction, working with log-ratio transformation such as additive log-ratio (alr), centered log-ratio (clr) and isometric log-ratio (ilr) allows to overcome some of the issues related to the complexity of geochemical data, helping to highlight the relative magnitudes and variations of the components of a composition rather than their absolute values (Buccianti and Magli, 2011). In this study, due to its orthonormal character, ilr was applied on the datasets and compared to raw and log-normal data to show how it allows normalizing the data distribution and its closure effects of geochemical data prior to statistical analyses (Figs. 2–4).

This log-ratio transformation was applied on data taking into account the compositional vectors of n parts partitioned into groups of parts presenting a certain affinity (Egozcue et al., 2003; Filzmoser and Hron, 2008; Filzmoser et al., 2009a, 2009b; Thiombane et al., 2018a).

To better visualize the element distributions and possible natural and anthropogenic behaviour of the variables into the survey area, a compositional biplot was created. This is a powerful statistical tool that displays both samples and the variables of a data matrix in terms of the resulting scores and loading (Gabriel, 1971). Thus, the scores represent the structure of the compositional data into a Euclidian space based on variance and covariance matrix; moreover, they display the association structure of the dataset. The biplots present rays (or vectors) defined from the center of the plot, where their length is proportional to the amount of explained variance (communality) of the variables it

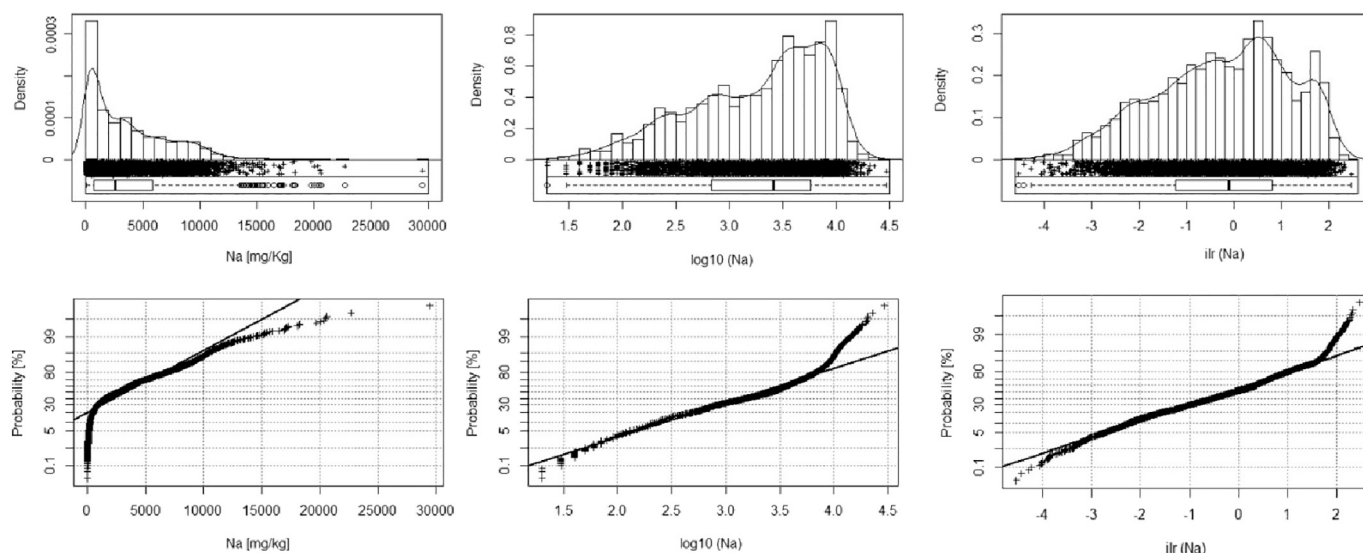


Fig. 2. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of Na of the raw data, log-transformed data, and ilr transformed data.

represents. The interpretation of the graphic depends on the loading (rays) structures and in more details on the approximate links between rays and samples, the distances between vertexes and their directions (Otero et al., 2005). For a full description of compositional biplots and an appreciation of their utility, several examples are available in the literature (e.g., Pison et al., 2003; Maronna et al., 2006; Filzmoser and Hron, 2008; Filzmoser et al., 2009a, 2009b; Hron et al., 2010).

When, however, biplots are used with raw data, these can be substantially influenced by the occurrence of outliers which can mislead the compositional nature of the data matrix and affect the principal components when interpreting results (Aitchison, 1986; Filzmoser et al., 2009a, 2009b). For these and others reasons log-transformed data are recommended to be used in multivariate analysis, and strengthened compositional biplots (Egozcue et al., 2003; Filzmoser et al., 2009a; Hron et al., 2010). Taking into account the singularity of the clr transformed data (Aitchison, 1986), these should be computed in orthonormal coordinates such as ilr transformed data, and back-transformed to the clr space for further interpretation. This back transformation allows preserving the linear relationship between clr and ilr

coordinates (Egozcue et al., 2003). Furthermore, the application of the minimum covariance determinant (MCD) estimator (Rousseeuw and Van Driessen, 1999) allows displaying the observations to the smallest determinant of their sample covariance matrix which tend to hold the variables into a normal distribution. In this study, a classical compositional biplot (CCB) and a robust compositional biplot (RCB) were used to identify the relationships between variables using compositional raw data and log-transformed data, respectively (Fig. 5).

From the total of 53 elements analyzed for the soils, only eighteen elements were considered to test this approach, with the aim of better representing the correlation between variables and investigate more robustly their main sources in the study area. Their main descriptive statistics are shown in Table 1.

The number of elements was reduced to 18 variables based on 3 main criteria: 1) the removal of elements with > 40% of values below the detection limit (LOD), 2) choosing arbitrary mostly two representative Rare Earth elements which are geochemically congruent and 3) choosing elements with a communality of extraction higher than 0.5 (50%) and/or common variances < 0.5 (e.g. Reimann et al., 2002).

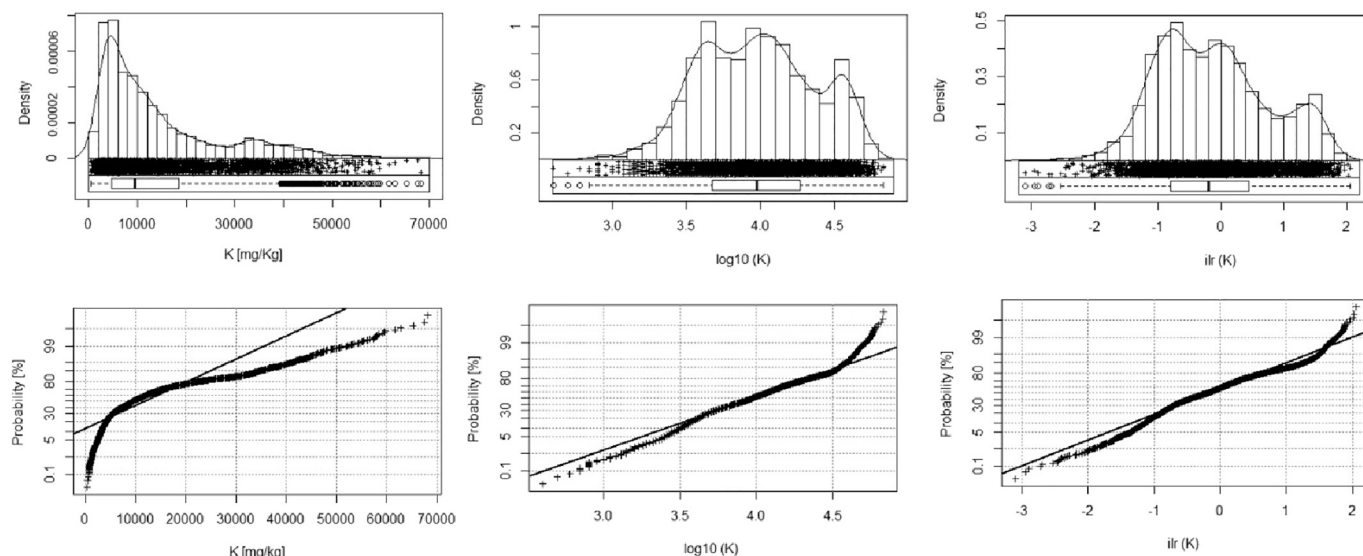


Fig. 3. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of K through the raw data, log-transformed data, and ilr transformed data.

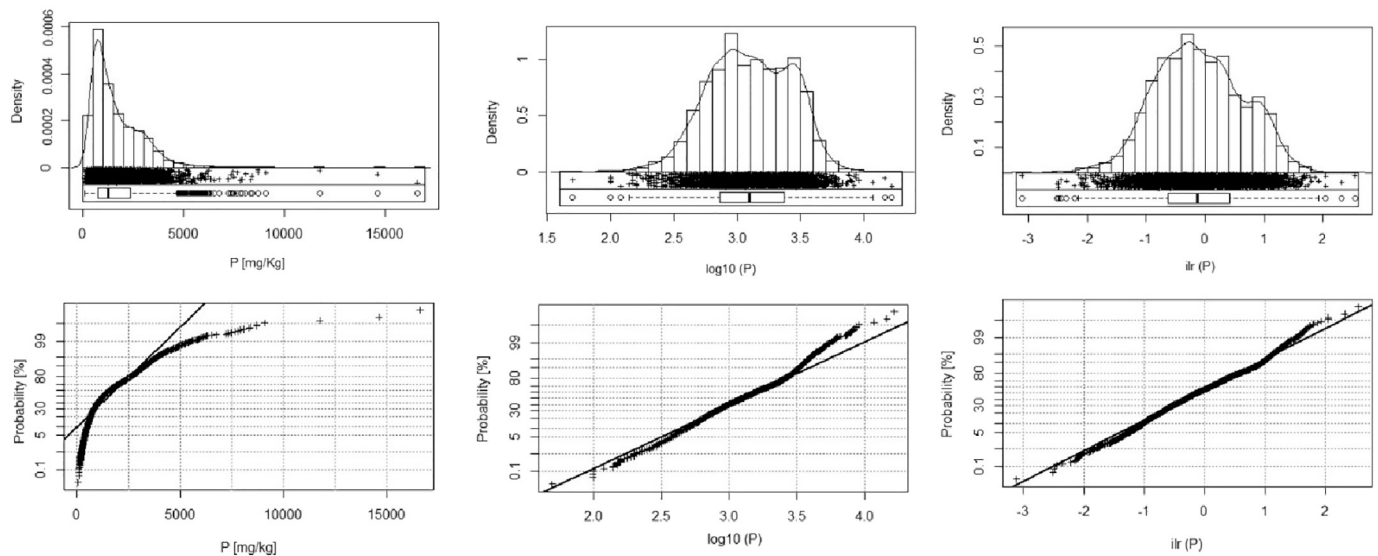


Fig. 4. Edaplot (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) and CP plot of P of the raw data, log-transformed data, and ilr transformed data.

Based on the robust biplot, sequential binary partition was implemented using the same 18 variables by dividing them into specific groups of non-overlapping elements (Table 2).

Balances are particular ilr-coordinates (isometric-logratio) having orthonormal bases which can be interpreted in the D-1 (D: dimension) real space as ratios of elemental associations (Egozcue et al., 2003). Balances can be calculated using the following formula:

$$Z_i = \sqrt{\frac{rs}{r+s}} \ln \frac{(\prod_{+} x_j)^{1/r}}{(\prod_{-} x_k)^{1/s}} \text{ for } i = 1, \dots, D-1, \quad (2)$$

where the products \prod_{+} and \prod_{-} only include parts coded with + and –, and r and s are the numbers of positive and negative signs (parts) in the i-th order partition, respectively (Egozcue and Pawłowsky-Glahn, 2005). From the established sequential binary partition and Eq. (2), Z_1

(Al/P) and Z_2 (Na/K) were calculated and ilr coordinates displayed through geospatial mapping. Balances can be interpreted considering three major cases: 1) positive balance when parts (variables) in the numerator have higher dominance with respect to parts involved in the denominator; 2) negative balance when parts involved in the numerator have lower dominance than those in the denominator (negative balance); 3) nearly zero balance when the dominance of the two groups of parts is similar. The higher or lower the positive or negative balance, respectively, the dominance of one group of parts is more pronounced. Balances were back-transformed based on the sequential binary partition matrix and the bijection between the original space of the parts and that of the log-ratios (Egozcue et al., 2003; Egozcue and Pawłowsky-Glahn, 2005; Olea et al., 2018). The back-transformed results in the original part space for Al, P, Na and K elements concentrations were

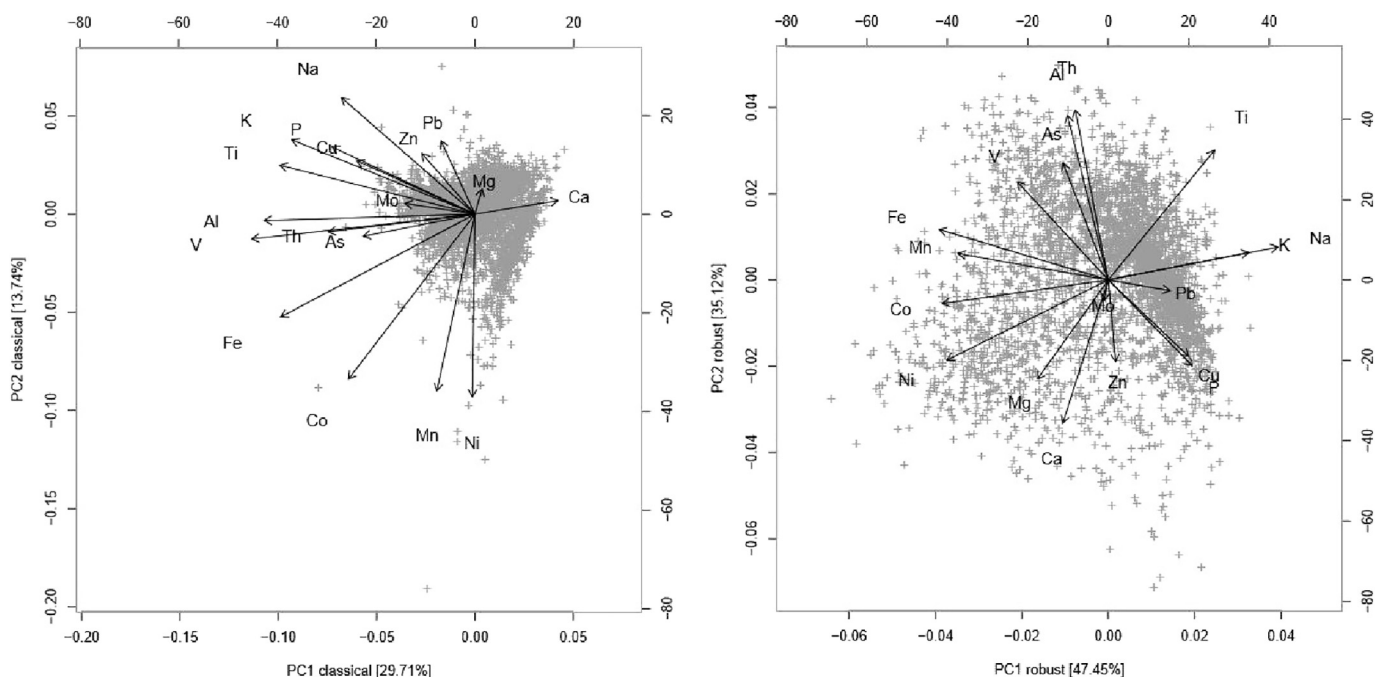


Fig. 5. Biplots for first and second principal components of factor analysis for raw data (classical, left plot) and ilr coordinates clr back-transformed (robust, right plot) of the survey area (Campania region, Southern Italy).

Table 1

Descriptive statistic of 3669 topsoils samples from the Campania region, Southern of Italy. RMS and Std. Deviation are the root mean square and standard deviation, respectively.

Elements	Unit	Minimum	Maximum	Mean	Median	RMS	Std. deviation	Skewness	Kurtosis
Al	mg/kg	2100	94,700	40,584	41,600	44,014	17,036	0.01	−0.75
As	mg/kg	0.6	163	12.6	12.1	14.6	7.4	6.1	87.8
Ca	mg/kg	800	295,200	35,495	22,300	51,036	36,675	2.38	7.1
Co	mg/kg	0.5	79	10.7	10.3	11.7	5.1	1.6	12.3
Cu	mg/kg	2.5	2394	109.3	62.2	191	156.8	5.5	44.5
Fe	mg/kg	1600	154,600	25,031	2510	26,282	8012	1.19	18.2
K	mg/kg	400	68,200	14,008	9500	18,755	12,472	1.39	1.2
Mg	mg/kg	700	104,600	7347	5800	10,483	7479	5.06	35.8
Mn	mg/kg	77	7975	863	779	970.6	443.4	5.44	55.7
Mo	mg/kg	0.06	62	1.5	1.3	2.2	1.6	18.9	637
Na	mg/kg	20	29,490	3667	2600	5129	3587	1.23	1.68
Ni	mg/kg	0.5	100	16.2	14.8	19.6	10.9	2.42	9.97
P	mg/kg	50	16,620	1641	1250	2063	1250	2.20	12.1
Pb	mg/kg	3.1	2052	73.8	54.2	117.7	91.7	7.7	100.2
Th	mg/kg	0.3	60	12.8	12.4	14.5	6.7	1.1	3.5
Ti	mg/kg	5	3270	1159	1240	1314	618.7	−0.23	−0.65
V	mg/kg	5	224	66.9	62	73.5	30.3	0.51	−0.29
Zn	mg/kg	11.4	3210	119	91	168.1	118.5	9.2	164.7

Table 2

Sequential binary partition table of the 18 investigated variables and the obtained balances (Z_1 – Z_{17}). Parts coded with + and − are the elemental associations involved in the calculation of the i-th order partition, respectively.

Balances	Ti	Th	As	V	Al	P	Na	K	Mo	Cu	Zn	Pb	Ca	Mg	Fe	Mn	Co	Ni
Z_1					+	−												
Z_2							+	−										
Z_3					+	+	−	−										
Z_4	+	+	+	−	−													
Z_5				+	−													
Z_6	+	+	−															
Z_7	+	−																
Z_8						+	+	+	+	+	+	+	−	−				
Z_9									+	−								
Z_{10}											+	−						
Z_{11}								+	−									
Z_{12}													+	+	−	−	−	−
Z_{13}													+	−				
Z_{14}															+	+	−	−
Z_{15}															+	−		
Z_{16}																	+	−
Z_{17}																		

computed before applying geostatistical computations.

2.4. Interpolated and background/baseline maps

Geographical Information Systems (GIS) and technology was used to map and display data distribution and characterize the footprint, possible main sources, and the behaviour of the elements considered. For this study, one of the aims was to determine the background concentration of major elements Al, Na, K, and P in the Campania soils. ArcGIS (ESRI, 2012) and GeoDAS (Cheng et al., 2001) were used as the main GIS tools. In particular, GeoDAS™ was used to produce interpolated geochemical maps by means of the multifractal inverse distance weighted (MIDW) algorithm (Lima et al., 2003a). In previous geochemical studies of the Campania region (De Vivo et al., 2001; Lima et al., 2003a; Cicchella et al., 2005; Albanese et al., 2007), the MIDW was chosen as an interpolation method as it preserves high frequency information (anomalies), while taking into account both spatial associations and local singularity in geochemical data (Cheng, 1999). The concentration–area (C–A) fractal method (Cheng et al., 1994) was applied to set the concentration intervals of the interpolated surfaces generated by the MIDW method, and ArcGIS™ software was used for the graphical presentation of the results (Fig. 6).

Different tools are used to determine the background and baseline

concentration of elements (EPA, 2001; Reimann et al., 2005; APAT-ISS, 2006; Tarvainen and Jarva, 2011; Cave et al., 2012; Ander et al., 2013). They are called “traditional approaches” due to the fact that most of them do not take into account both spatial association and the data distribution local singularity. In this study, maps showing geochemical background/baseline concentrations have been obtained using the S–A (spectrum–areas) method which preserves high frequency information. The S–A method is a fractal filtering technique, based on a Fourier spectral analysis (Cheng, 1999; Cheng et al., 2001), and is used to separate anomalies from background values starting from a geochemical interpolated concentrations map. It also uses both frequency and spatial information for geochemical map and image processing. Fourier transformation can convert geochemical values into a frequency domain in which different patterns of frequencies can be identified. The signals with certain ranges of frequencies can be converted back to the spatial domain by inverse Fourier transformation (Zuo et al., 2015; Zuo and Wang, 2016). The interpolated maps generated from geochemical data have been transformed into the frequency domain in which a spatial concentration–area fractal method has been applied to distinguish the patterns on the basis of the power–spectrum distribution. A log–log plot (Fig. 8; Fig. 10 under plots) was used to show the relationship between the area and the power spectrum values on the Fourier transformed map of the power spectrum. The values on the

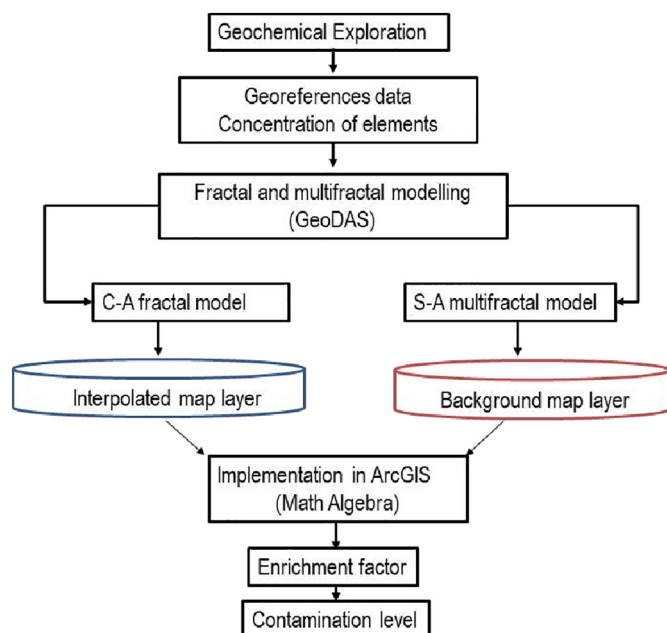


Fig. 6. Flow chart of data processing for contamination degree modelling using GIS environment.

log–log plot were modelled by fitting straight lines using least squares. Distinct classes can be generated, such as lower, intermediate, and high power spectrum values approximately corresponding to baseline values, anomalies, and noise of geochemical values in the spatial domain (Fig. 9; Fig. 11, under plots), respectively.

2.5. Enrichment factor

The Enrichment Factor (EF) approach, which was historically introduced to identify the level of (economically viable) mineralisation and origin of elements in atmosphere, precipitation or seawater (Goldberg, 1972; Chester and Stoner, 1973; Peirson et al., 1974; Duce et al., 1975; Rahn, 1976; Buat-Ménard and Chesselet, 1979) was used in this study to ascertain soil contamination on a long term scale (see for example Hakanson, 1980; Sutherland et al., 2000; Abraham and Parker, 2008; Wu et al., 2011; Saeedi et al., 2012). EF is computed using the equation described below (Eq. (2)) which was first introduced by Chester and Stoner (1973):

$$EF = \frac{\left(\frac{C_x}{C_{ref}}\right)_{sample}}{\left(\frac{C_x}{C_{ref}}\right)_{background}}, \quad (3)$$

where C_x is the concentration of the element under consideration and C_{ref} is the concentration of a reference element. Here, the reference element is an element that is particularly stable in soil. In fact, the stability of the element is demonstrated by a vertical immobility and/or his chemical stability (non-degradability) (Reimann et al., 2008). Aluminium, Sc, Zr and Ti are the main elements considered in the literature to be stable, and naturally occurring in soils. In this study, the choice of the most stable element in EF computation was based on a robust statistical estimation called coefficient of variation (CV) which allows a more extensive interpretation of the variability of distribution of reference elements (Al, Zr, Sc, and Ti) using the equation:

$$CV = \frac{MAD}{MD} \times 100\%, \quad (4)$$

where CV displays the variability of distribution in percentage (%), MAD corresponds to the median absolute deviation that is the median value (50th percentile) of the deviations of all concentrations from the

Table 3

Variability Test of Al, Sc, Ti and Zr elements; MD is the Median, MAD corresponds to median absolute deviation, $MAD = \text{median}\{|x_i - \text{median}(x_i)|\}$, that is the median value (50th percentile) of the deviations of all individual x_i values (concentrations) from the median value (concentration). CV = coefficient of variation.

Stat. parameters	Al (ppm)	Sc (ppm)	Ti (ppm)	Zr (ppm)
MD	41,600	2.3	896	5.1
MAD	10,900	1.01	240	1.7
CV (%)	26.2	43.8	26.7	34.2

median value of concentration and MD is the median concentration. This is a robust, nonparametric estimate that is not affected by the presence of outliers (Reimann and de Caritat, 2005). The lower the CV value, the more the element is stable (Table 3).

This study intended to investigate the most effective EF calculation by comparing the use of the reference element in continental crust (Martin and Whitfield, 1983; Peirson et al., 1974; Taylor and McLennan, 1995; Wedepohl, 1995; Loska et al., 1997) versus the use of the reference element of the local background area as advised by Reimann and De Caritat (2005) and Sutherland et al. (2000). The variation of P EFs in the study area was then displayed by means of interpolated maps where range of EF scores were calculated taking into account the contamination factor in accordance with Sutherland et al. (2000) (Fig. 12, see legend).

3. Results and discussion

3.1. Univariate and multivariate analysis

Results for Na, K and P elemental distribution have been presented by combining Edaplots (top) and CP plots (bottom) with three different data type: raw data concentration (left), log-normal data (middle) and ilr transformed data (right) (Figs. 2, 3 and 4).

The Edaplots for Na (Fig. 2) show different ‘shapes’ depending on their type of data: the raw data distribution is right-skewed while the log-normal data is left-skewed. This highlights how both raw and log-normal data representation do not match well the real data distribution compared to the ilr transformation. By using the ilr transformed data, the distribution (as shown by histogram and density plot) tends to a normal data distribution. A similar result can be observed for K and P data distribution (in Figs. 3 and 4). The strength of ilr transformation in data distribution is shown in the CP plot, which displays the cumulative curve distribution where the straight-line symbolizes the most adequate model of a normal data distribution; by using the ilr transformed data, the elemental distribution fits very closely the straight line compared to raw and log-normal data, which are affected by the occurrence of outliers.

The compositional biplot (Fig. 5), based on principal component analysis, displays the correlation and relationship between 18 analytical variables in 3669 sample points, from which the first two principal components were extracted. The principal components are presented in a compositional biplot using raw data (Fig. 5, left) and ilr coordinates clr back-transformed (Fig. 5, right).

The total variance of initial raw data biplot (classical biplot) explains 43.45%, where the first principal component (PC1) accounts for 29.71% and the second principal component (PC2) accounts for 13.74% (Fig. 5, left). On the other hand, the robust biplot based on ilr coordinates and clr back transformed produced significantly different results (Fig. 5, right) with PC1 explaining 47.45% and PC2 explaining 35.12% of the compositional variability.

By taking into account the direction and angles formed between the vectors, and the proximity of the rays, it is possible to identify the presence of four groups of elements, which are most likely related to the geogenic features and/or the main human activities in the study areas.

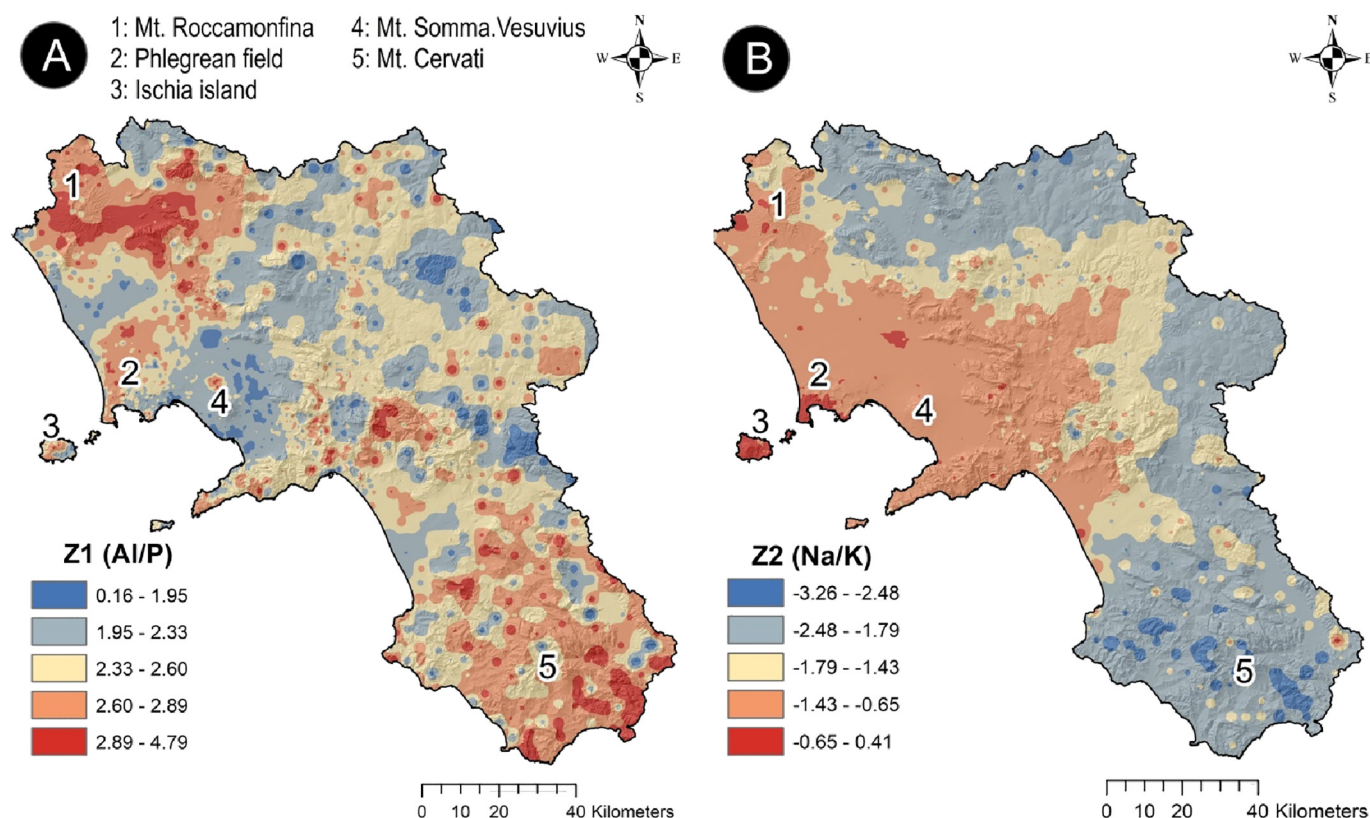


Fig. 7. (A) The interpolated Z_1 map. Note the red and blue colours highlight higher and lower proportion of Al and P, respectively. (B) The interpolated Z_2 map. The red and blue colours highlight higher and lower proportion of Na and K, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The main groups highlighted by the biplots are:

- ✓ Fe, Mn, Co and Ni, a group of element association characterized by their rays closed to one-another and tending to the same direction (classical biplot). This group of element is marked by high communalities (length of the rays) of the vectors. This may be expected given that this association is strongly related to the adsorption and coprecipitation effects operated by Fe and Mn oxides and hydroxides occurring mostly in the sedimentary deposits such as marl-sandstone, conglomerate and silico-clastic flysch deposits outcropping in the surveyed areas (Cicchella et al., 2005; Albanese et al., 2007; Buccianti et al., 2015).
- ✓ Al, Th, As, V, Ti, and Mo form an association of elements where Al, Ti and V dominate the groups with the highest length of their vertexes whilst Mo has a relatively lower communality (classical biplot). This behaviour is possibly related to the fact that these elements are mostly immobile during weathering phenomena of the parental rocks and mostly remaining in the residual fraction of soils. This group could therefore be directly related to the parental geology of the surveyed areas which are dominated by the influence of pyroclastic deposits from different eruptions of nearby volcanoes such as Roccamonfina, Vesuvius, Campi Flegrei (De Vivo et al., 2010).
- ✓ Na, K, P, Cu, Pb and Zn elemental association is dominated by a high communality of Na and P as well as the vicinity of their rays (classical biplot). This elemental association probably reflects the potassic and ultrapotassic rock formations that occur throughout the majority of the slope of Naples and Benevento areas, associated to the lava and pyroclastic volcanic activity of Mt. Somma-Vesuvius and Roccamonfina (Lima et al., 2003b; Albanese et al., 2013). Zinc and Pb display short vectors which are poorly characterized and seem to be only partially correlated to the others element of this

association. These two elements may be related to anthropogenic activities such fossil fuel combustion, as well as industrial and vehicular emissions release (Cicchella et al., 2005; De Vivo et al., 2016).

- ✓ Ca and Mg appear to be correlated and present both a short length of their rays (classical biplot). The communality of Ca is larger and seems to be independent of all others elements due to the fact that the angles formed are $> 90^\circ$ compared to Mg. This confirms the high correlation between Ca and Mg, which might be possibly related to the limestones and dolostones of the Mt. Picentini, Mt. Lattari and Mt. Cervati.
- ✓ On a closer observation, the elemental association Na, K, P, Cu, Pb and Zn could be reduced to two main subgroups, where the variables K and Na are strongly overlapping on to each other (robust biplot). This highlights the high correlation between these two elements occurring mostly in potassic and ultrapotassic rock formations throughout the surveyed areas. Phosphorous and Cu are also highly correlated both with high communalities, where P seems to be independent of the Na/K vertexes, forming almost an angle of 90° (robust biplot). Interestingly, Cu seems to be poorly correlated to a geogenic origin, as the direction of its ray compared to those of the group of sulfide elements (e.g. Co and Ni) have an angle up to 90° . This may signify that Cu is independent from other naturally occurring sulfide elements, whilst it seems to be highly correlated to P in most soil of the surveyed area. One potential explanation is that P and Cu may be related to agriculture activities, with large areas cultivated as vineyards, where the use of pesticides and phosphate fertilizers is very high (Cicchella et al., 2005).
- ✓ For the Al, Th, As, V, Ti, and Mo association, it is observed the dissociation of Ti variable with a high communality of the ray (robust biplot). Titanium is considered as an immobile and stable element due to low mobility and is mostly found in volcanic materials

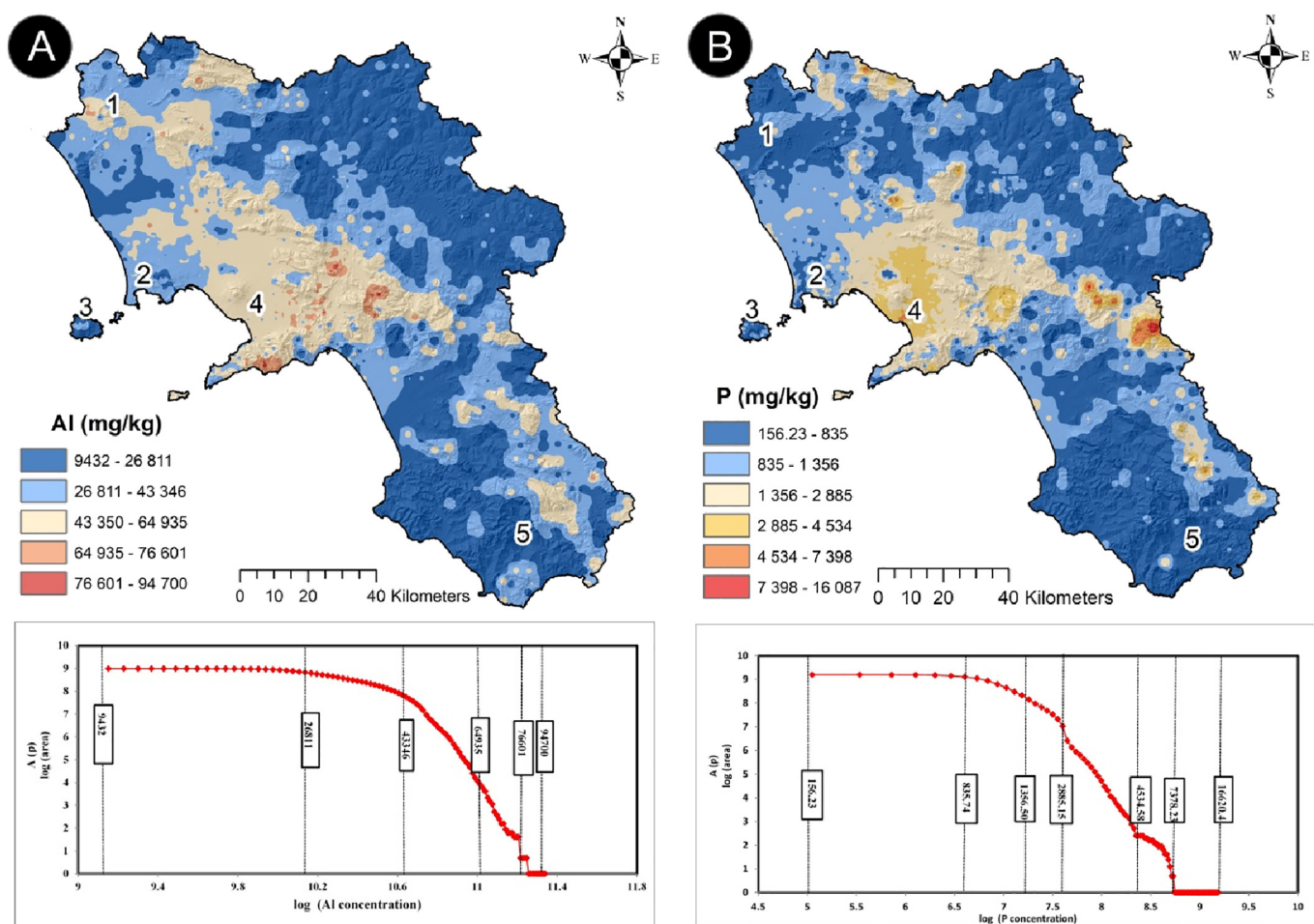


Fig. 8. Interpolated maps of Al (A) and P (B) elemental distribution in the survey area; ranges of concentration are based on the C-A fractal plot held below.

(Egli et al., 2008).

3.2. Geochemical elemental distribution in the survey areas

Based on the robust biplot, 18 elements have been chosen to perform sequential binary partition and obtain balances (specific ilr-coordinates) (Table 2). In this section, Balances Z_1 (Al/P) and Z_2 (Na/K) will be displayed to reveal the data for the elements of main interest (Al, P, Na and K).

The first balance Z_1 (Al/P) map, ranging from 2.89 to 4.79 reveals a higher proportion of Al in correspondence to large volcanic complexes like Mt. Roccamonfina and Phlegraean fields (Fig. 7A). In addition, high proportion of Al is also highlighted in correspondence of part of the Lattari range, along the Apennines and in patches at the southern part of our study area. Scheib et al. (2014) highlighted that the Mt. Roccamonfina volcano is characterized by pyroclastics rocks with high level of elements such Al, Th and Ti, as well as in the Campanian Ignimbrites in the Apennines (De Vivo et al., 2010).

In contrast, the higher proportions of P (Fig. 7A) in correspondence to lower values of coordinate (ranging from 0.6 to 1.95) are found around Mt. Somma-Vesuvius, and in several areas of the eastern region of our study area, where large agricultural fields (e.g., vineyards and orchards) are located.

The second balance map Z_2 (Na/K) (ranging from -0.65 to 0.41) shows the dominance of Na and K in the study area (Fig. 7B). In fact, higher proportion of Na can be observed in correspondence to Mt. Roccamonfina, Phlegraean fields, and Ischia Island. Sodium may be related to the potassic and ultrapotassic rocks and volcano-sedimentary deposits from major sector collapse of volcanoes in the study area

(Scheib et al., 2014; De Vivo et al., 2016). In contrast, the higher abundance of K corresponding to lower balances (ranging from -3.26 to -2.48), can be observed in the southern part of our study area. In fact, Thiombane et al. (2018a) showed a high enrichment of K in silico-clastic deposits dominated by flysch series in southern part of our study area. Furthermore, pyroclastic rocks from different eruptions of nearby volcanoes (Roccamonfina, Vesuvius, Phlegraean Fields - De Vivo et al., 2010; Buccianti et al., 2015; Mt. Vulture and Aeolian islands - Peccerillo, 2005; Scheib et al., 2014) are found in this area. The back-transformation of balances based on the partition matrix helped to display results in the original part space giving the same values of Al, P, Na and P elements concentration because of the bijection between the original space of the parts and that of the ilr-transformation (Egozcue et al., 2003; Olea et al., 2018).

The concentration of Al, ranging from 2344 to 94,334 mg/kg with a mean value of 32,918 mg/kg, was separated into five ranges according to C-A fractal plot (Fig. 8A, plot below).

The lowest concentration values roughly ranging from 2344 to 30,000 mg/kg, are found in the north-eastern and south-western part of the study area in correspondence with the Apennine chain and the Cervati Mt., respectively. The highest concentrations (up to 57,000 mg/kg) are found in soils on the slope of the volcanoes (Mt. Somma-Vesuvius and Roccamonfina), surrounding the Mt. Matese and the Mt. Lattari (Fig. 8A). The elevated concentration of Al in soils surrounding the volcanoes is possibly related to the parental pyroclastics which subsequently formed soils (De Vivo et al., 2016). On the other hand, in the Mt. Matese and Mts. Lattari, Al concentrations could result from the occurrence of several imbrications of bauxite minerals which were exploited in the first part of the 20th century (Mondillo et al., 2011).

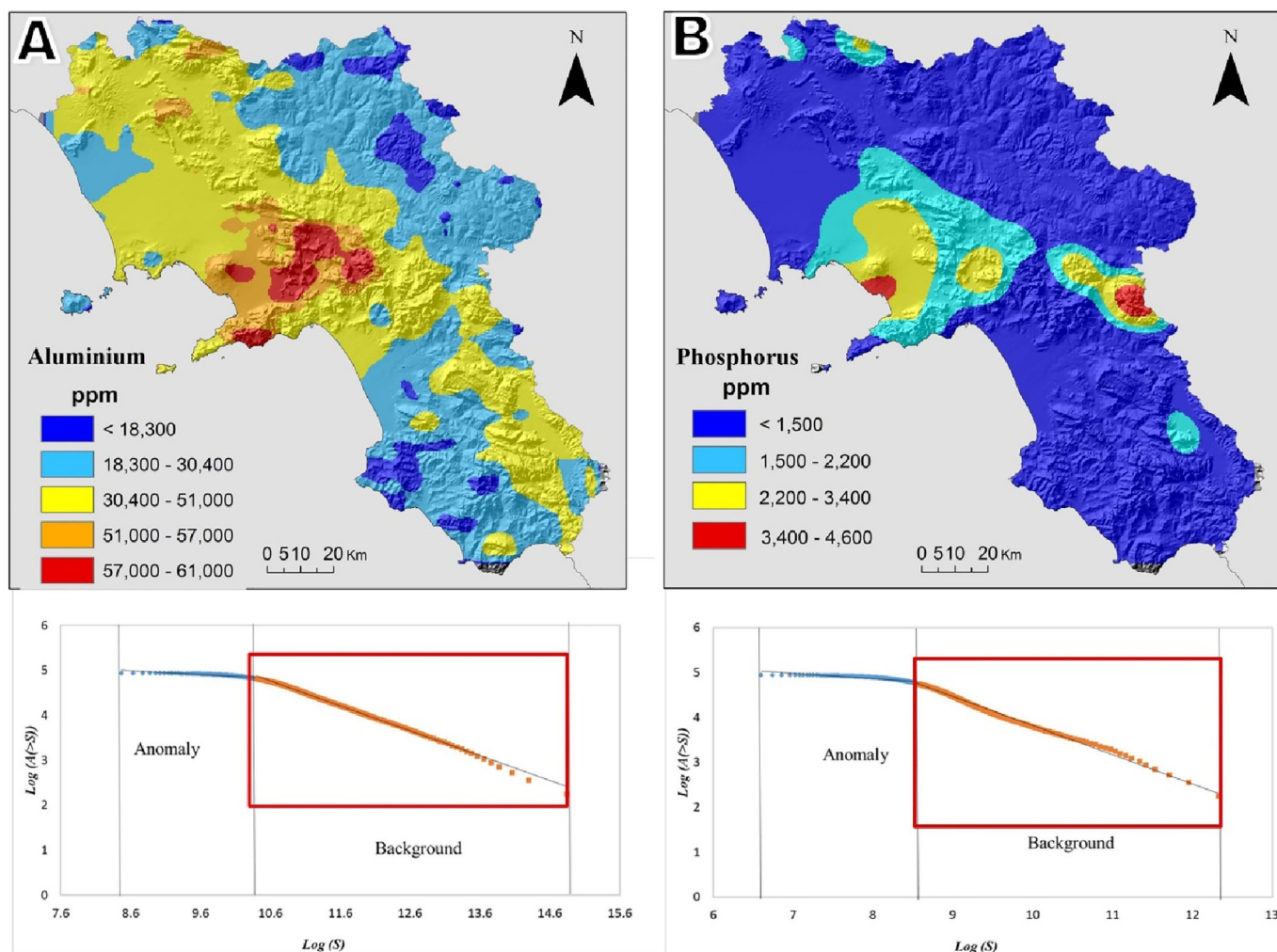


Fig. 9. Background/baseline maps of Al (A) and P (B) elemental concentration in the survey area.

The background/baseline map of Al (Fig. 9A) shows the highest concentration ranging from 51,000 to 61,000 mg/kg in correspondence with soils around the Roccamonfina and Vesuvius volcanoes.

Phosphorus content in Campania soils ranged from 156 to 16,087 mg/kg (Fig. 8B) with a mean value of 1011 mg/kg, which corresponds to the mean level of the European soil (Tóth et al., 2014). Average concentrations (ranging from 64,935 to 75,000 mg/kg) increased significantly near Mt. Somma-Vesuvius and the highest values (up to 75,000 mg/kg) were found in Vitulano municipality (Benevento Province), and in the nearby Lioni and Laviano districts (Avellino Province). In particular, the corresponding underlying geology does not seem to be able to account for this P elemental anomaly: the soils of Vitulano district are mostly from sandstone, flysch deposits and limestone whereas Lioni and Laviano districts soils are mostly from limestones. The high P concentration might be possibly related to anthropogenic activities related to the large fertilizers use in agriculture. Part of the Vitulano municipality lies in fact in an area with intensive vineyard occurrence. The background/baseline P distribution (Fig. 9B) is characterized by > 95% of the survey area with P values < 1500 mg/kg. This range of concentration is usually characteristic of siliciclastic, limestone and dolostone deposits/geology. Greater concentrations ranging from 2200 to 4600 mg/kg were found on the slopes of the Mt. Somma-Vesuvius, and Lioni and Laviano districts, possibly linked to geogenic and anthropogenic sources, respectively.

Sodium and K concentrations ranged from 20 to 17,592 mg/kg with a mean concentration of 1265 mg/kg, and from 804 to 63,850 mg/kg with a mean concentration of 6852 mg/kg, respectively (Table 1). The

highest values of Na (up to 10,900 mg/kg) were found in soil samples on the slopes of the Somma-Vesuvius and Roccamonfina volcanoes, in the Phlegrean Fields and Ischia Island whereas those for K (up to 41,522 mg/kg) were found mostly on soils surrounding the Mt. Somma-Vesuvius (Fig. 10A and B). These values could reliably be attributed to the occurrence of volcanic rocks and related soils in the surveyed areas. These formations are dominated by potassic and ultrapotassic lavas and pyroclastic materials. The interpolated maps reflect hence the concentrations of Na and K elements in such rocks and pyroclastics formations linked to Quaternary volcanic activities (Peccerillo, 2005; Albanese et al., 2013; Bucciatti et al., 2015). In general, the soils of the Vesuvian area, formed from a more recent volcanic activity (Joron et al., 1987; De Vivo et al., 2003; Lima et al., 2003b) are much richer in Na and K than the soils on the much older Roccamonfina volcano. This is due to the fact that Na and K are relatively mobile and easily leached elements in the surficial environment.

The background maps of the Na and K (Fig. 11A and B) show low concentration ranging from 20.1 to 3200 mg/kg and from 804 to 9100 mg/kg, respectively. These relatively low ranges of concentration of Na and K were found to correspond to the same lithologies. At regional level, the highest concentration of Na and K was found in soil on the slopes of the Roccamonfina and Mt. Somma-Vesuvius, ranging from 3200 to 9700 mg/kg and from 9100 to 46,900 mg/kg, respectively. This allows delimiting two ranges of background concentration of Na and K related to the local geology (Table 4).

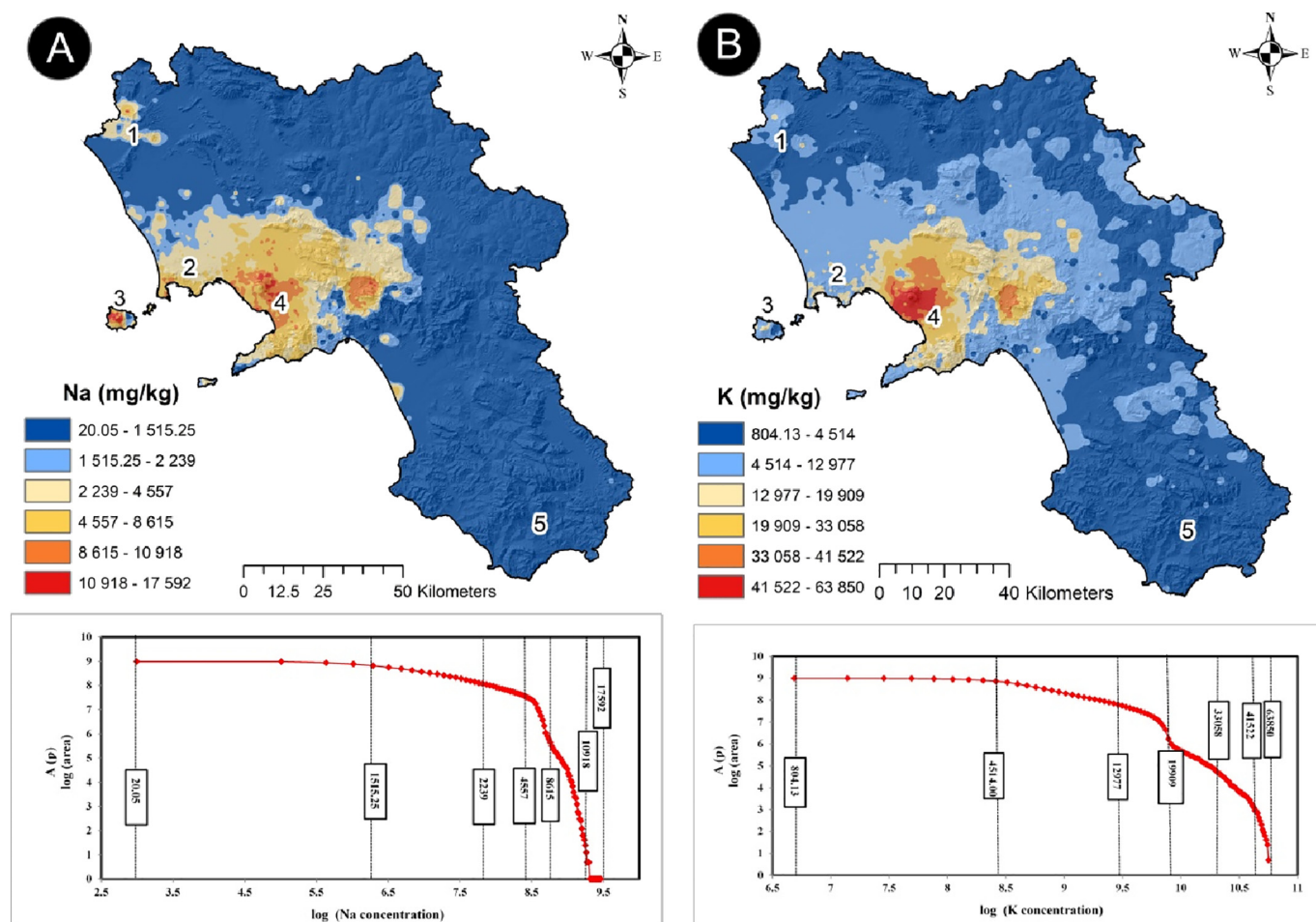


Fig. 10. Interpolated maps of Na (A) and K (B) elemental distribution in the survey area; ranges of concentration are based on the C-A fractal plot held below.

3.3. Enrichment factor or phosphorus degree of contamination

The variability coefficient (CV) of the reference elements was measured as: 26.2% (Al), 43.8% (Sc), 26.7% (Ti) and 34.2% (Zr - Table 3). Aluminium displayed the lowest value of CV, followed closely by Ti, Zr and Sc. Al and Ti are mostly related to processes forming and presence of oxides (Al_2O_3 and TiO_2) in clastic materials and are not easily affected by weathering processes. These results prove that Al remain the most stable element and was therefore chosen as reference variable to determine the P enrichment factor for this study too.

The enrichment factor scores were calculated for P using both the Al elemental concentration in continental crust (Fig. 12A), and the local Al background concentration in the survey areas (Fig. 12B).

The reference with the continental crust value (Fig. 12A) showed a lower EF score ($\text{EF} < 2$) in the northern and southern parts of the region, with medium EF scores (ranging from 2 to 4) comprising > 50% of the study area. Higher EF scores (ranging from 20 to 40), corresponding to anomalous enrichments, were found in soil on the slopes of the Mt. Somma-Vesuvius and in the Lioni and Laviano districts. Vesuvian areas (Fig. 9) displayed higher concentration of P which can though be related to the underlying background concentration, where the enrichment can be explained by geogenic (volcanic) source. Lioni and Laviano districts lie in areas characterized by limestone-derived soils; a high P enrichment factor could therefore be linked to anthropogenic activities. The highest range of P enrichment factor scores (> 40) is registered in the provincial territory of Benevento (particularly towards the Vitulano municipality). Soils in this area are derived from sandstone, flysch deposits and limestone, where intensive agriculture activities, such as vineyards and olive plantations, are practiced.

The extremely high P enrichment may be related to the use of phosphate fertilizers in agriculture practises in this area.

The calculation of P enrichment factors with local background concentrations (Fig. 12B) presented different distribution and intensity of EF scores in the study area. The Vesuvian areas presenting previously significant enrichment (EF ranging from 5 to 20, Fig. 12A) were now presenting a minimum enrichment instead ($\text{EF} < 2$, Fig. 12B). The Vesuvian area are characterized by a high P background concentration related to the underlying parental volcanic rocks which represent materials for the subsequently formed soils throughout this area. Furthermore, Al was found at lower concentration values compared to its value in the continental crust (Table 4). Indeed, EF calculation using the P and Al local background concentrations provides a smoother EF in relation to the Vesuvian area compared to their actual local variabilities.

A similar result was observed in the area surrounding Lioni, Laviano and Vitulano municipalities. EF calculated in the classic way (Fig. 12A) showed a very high (EF ranging 20 to 40) and extremely high enrichment ($\text{EF} > 40$) as opposed to the moderate (EF ranging from 2 to 5) and significant enrichment (EF ranging from 5 to 20) shown by EF calculated using the local background (Fig. 12B). Using the local background concentration of Al and P (Fig. 12B), the widespread enrichment created by calculation made using the continental crust values as a reference fades away, replaced by a much narrower enrichment that take into account the local variability of P and Al, and highlights the anthropogenic P inputs observed around the Lioni and Laviano districts, and even more significantly in the Vitulano municipality where it may be related to the use of phosphate fertilizers in agriculture activities. A continental crust concentration based EF 'hides' the

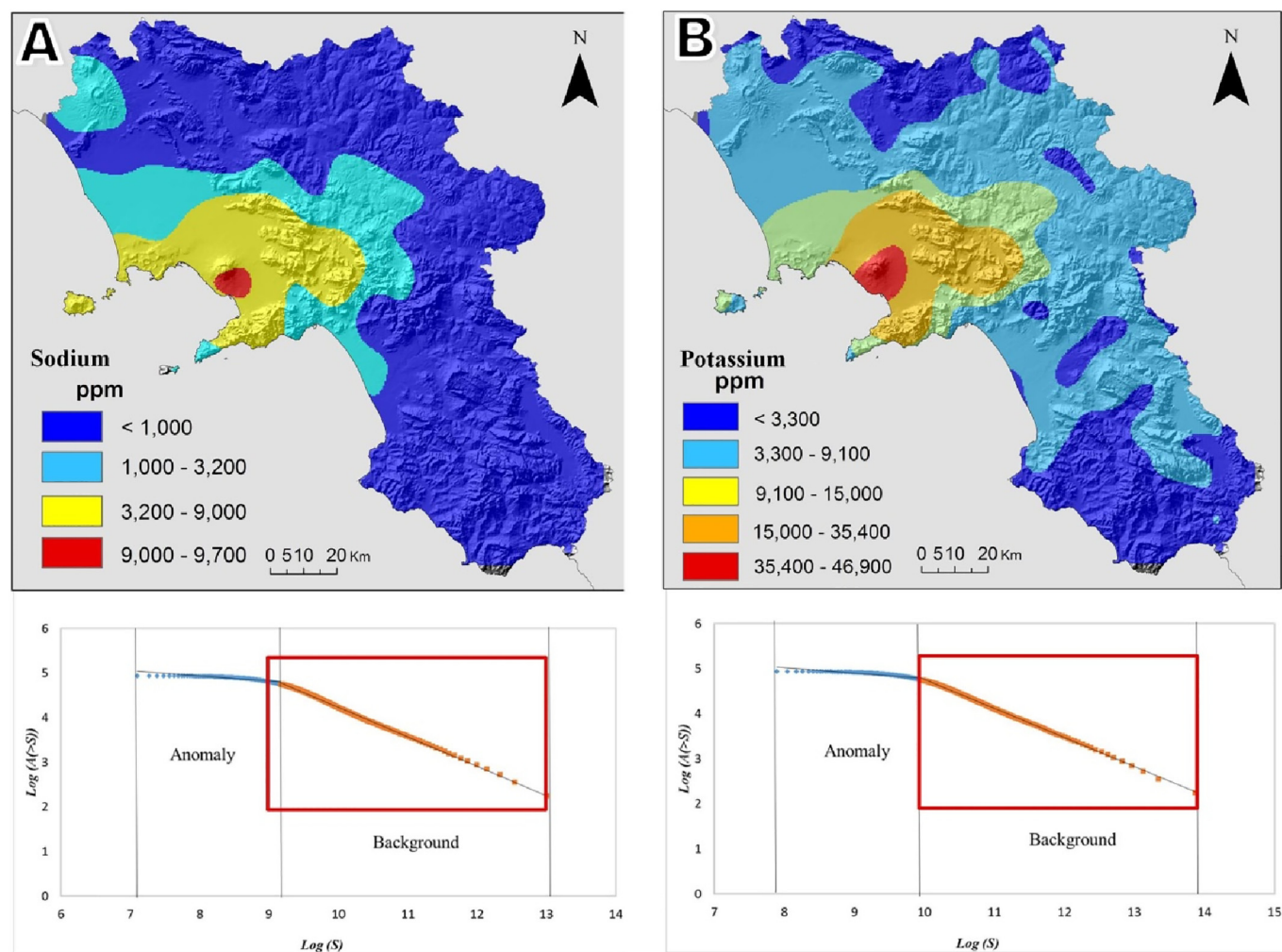


Fig. 11. Background/baseline maps of Na (A) and K (B) elemental concentration in the survey area.

Table 4

Geogenic background/Baseline value ranges for Al, P, Na and K elemental concentration in the continental crust (Wedepohl, 1995) and in soils of the Campania region according to the lithology.

Elements	Wedepohl, 1995 ^a		Local background in this study	
	Cont. crust	Siliciclastic deposits	Limestone and dolostone	Volcano-clastic deposits
Al (mg/kg)	150,500	9432–30,400	30,400–51,000	51,000–61,000
Na (mg/kg)	35,600	20.4–1000	20.4–1000	1000–9700
K (mg/kg)	31,900	804–9100	804–9100	9100–46,900
P (mg/kg)	1500	156–1500	156–1500	1500–4600

^a Average in continental crust. (Wedepohl, 1995).

anthropogenic input behind a more general enrichment due to the fact that it does not consider the local geological variability (Reimann and de Caritat, 2005; Albanese et al., 2013; Zuzolo et al., 2017). However, when taking into account the local concentration, the anthropogenic input can be more clearly distinguished from the geogenic enrichment factor. In other words, by using as a reference the local background, it is easier to isolate anthropogenic enrichment factors from the geogenic ones.

4. Conclusions

Evidence from this study showed that compositional data

transformations such as *ilr* transformation can help to solve the outlier artefacts and moves the composition sample space to the Euclidean Real space R^{-1} , which is failed by the classic statistical data transformation (Filzmoser and Hron, 2008).

The multivariate and integrated approach applied in this study on a multi-elemental geochemical dataset allowed to highlight the correlation between variables and helped identifying the main sources of elements in the surveyed region. Biplot based on transformed data were able to highlight the main geological features as well as the potential input of anthropogenic activities in the Campania region. In particular, four association complexes were identified:

- 1) Fe, Mn, Co and Ni: associated to the coprecipitation of Mn and Fe oxides-hydroxides in flysch and arenaceous material;
- 2) Ca and Mg: associated to limestones and dolostone outcrops occurrence of the Mt. Picentini, Mt. Lattari and Mt. Cervati limestones;
- 3) Al, Th, As, V, Ti, and Mo: associated to the pyroclastic coverings;
- 4) Cu and P: potentially associated to the agriculture activities through use of phosphate fertilizers.

The robust biplots allowed to display elements in a wider space and provided grounds for an enhanced data interpretation: the length of the rays, which is linked to the variability of *ilr* data *clr* back transformed as opposed to the variables themselves, permitted to emphasize their potential sources; the groups of elements highlighted by proximal rays, were used as evidence of either local geological features and/or anthropogenic activities.

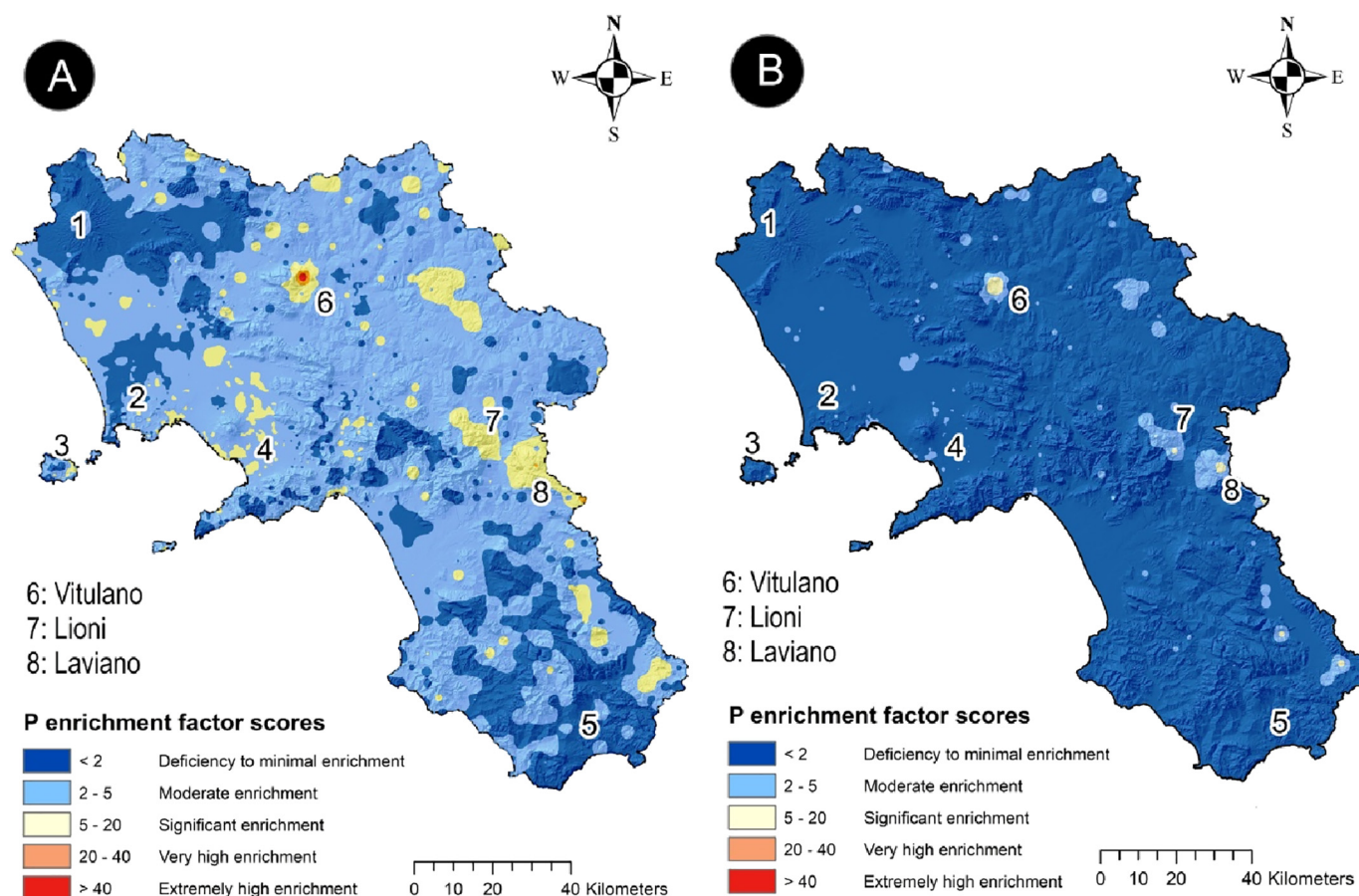


Fig. 12. Interpolated maps of the enrichment factor scores into the study areas; (A) map created using the continental crust reference values of Al and P (Wedepohl, 1995); (B) map is based on the local reference background concentrations of Al and P (this study) as reference elements in each pixel.

In addition, the interpolated maps by use of C-A fractal plot helped to distinguish element distributions related to their main sources, whereas using S-A multifractal plot allowed to display the background concentration Al, Na, K and P elements as reference values for Campania soils.

The maps of P Enrichment Factor scores using as reference the continental crust and the local background values showed:

- 1) A significant P enrichment (from 5 to 20) in soils on the slopes of the Mt. Somma-Vesuvius related to the underlying parental volcanic rocks (geology), where the geogenic component represent clearly the natural background. This enrichment disappeared when the local background reference of Al and P are used. In fact, due to the geogenic source of the P and Al in these areas, the EF tended to decrease when using the local background, confirming that those areas were not contaminated by human activities.
- 2) The highest range of P EF factor scores (> 40) were registered in the Benevento provincial territory (particularly in Vitulano district), Lioni and Laviano districts, with values decreasing two times when using as reference the local background values. Taking into account that the underlying geology of these areas could not influence these high P EFs, it is very likely that these values may have an anthropogenic source. The above provincial districts fall in a territory where intense agriculture activities are present, allowing to infer that such soils may be affected by use of phosphate fertilizers.

As a general observation, the findings from this study confirm the validity of using local background concentrations to better identify the 'real' degree of contamination as opposed to generalised continental crust values, which could emphasize 'spurious' enrichment due mainly

to natural local concentrations of elements.

From an applied point of view, the integrated approach applied here provided a more robust qualitative and quantitative evaluation, highlighting new and vital information on the distribution and patterns of key elements (Na, K, Al and P) in soils of the Campania region. The findings from this investigation strongly point towards highly desirable follow up studies in clearly identified and discrete areas displaying high P EFs: this would allow a more detailed and thorough assessment of P footprints, which could then be used for a comprehensive human health risk evaluation due to direct and indirect exposure. This is particularly important, when considering that, if P was linked to fertilizers source, these could include impurities such as Cu, As, Zn and Pb, well-known potentially toxic elements (PTEs).

Acknowledgments

This work was supported by grant from the Ministero dell'Istruzione, dell'Università e della Ricerca Scientifica - Industrial Research Project "BioPoliS" PON03PE_00107_01, funded in the frame of Operative National Programme Research and Competitiveness 2007–2013, D. D. Prot. N.713/Ric. 29/10/2010 and (EnerbioChem) PON01_01966, funded in the frame of the Operative National Programme Research and Competitiveness 2007–2013 D. D. Prot. n. 01/Ric. 18.1.2010. The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

References

- Abraham, G.M.S., Parker, P.J., 2008. Assessment of heavy metal enrichment factors and the degree of contamination in marine sediments from Tamaki Estuary, Auckland,

- New Zealand. *Environ. Monit. Assess.* 136 (1–3), 227–238.
- Aitchison, J., 1986. *The Statistical Analysis of Compositional Data*. Chapman & Hall, London (416 pp).
- Albanese, S., De Vivo, S., Lima, A., Cicchella, D., 2007. Geochemical background and baseline values of toxic elements in stream sediments of Campania region (Italy). *J. Geochem. Explor.* 93, 21–34.
- Albanese, S., Iavazzo, P., Adamo, P., Lima, A., De Vivo, B., 2013. Assessment of the environmental conditions of the Sarno River Basin (South Italy): a stream sediment approach. *Environ. Geochem. Health* 35, 283–297.
- Ander, E.L., Johnson, C.C., Cave, M.R., Palumbo-Roe, B., Nathanail, C.P., Lark, R.M., 2013. Methodology for the determination of normal background concentrations of contaminants in English soil. *Sci. Total Environ.* 454–455, 604–618. <https://doi.org/10.1016/j.scitotenv.2013.03.005>.
- APAT-ISS, 2006. Protocollo Operativo per la determinazione dei valori di fondo di metalli/metalloni nei suoli dei siti d'interesse nazionale. Revisione 0. Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici and Istituto Superiore di Sanità (in Italian).
- Bonardi, G., D'Argenio, D., Perrone, V., 1998. Carta geologica dell'Appennino meridionale. *Mem. Soc. Geol. Ital.* 41.
- Buat-Ménard, P., Chesselet, R., 1979. Variable influence of the atmospheric flux on the trace metal chemistry of oceanic suspended matter. *Earth Planet. Sci. Lett.* 42, 399–411.
- Buccianti, A., Magli, R., 2011. Metric concepts and implications in describing compositional changes for world river's water chemistry. *Comput. Geosci.* 37 (5), 670–676.
- Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V., 2006. Compositional Data Analysis in the Geosciences: From Theory to Practice.
- Buccianti, A., Pawlowsky Glahn, V., Egozcue, J.J., 2014. Variation diagrams to statistically model the behavior of geochemical variables: theory and applications. *J. Hydrogeol.* 519, 988–998.
- Buccianti, A., Lima, A., Albanese, S., Cannatelli, C., Esposito, R., De Vivo, B., 2015. Exploring topsoil geochemistry from the CoDA (Compositional Data Analysis) perspective: the multi-element data archive of the Campania Region (Southern Italy). *J. Geochem. Explor.* 159, 302–316.
- Buccianti, A., Lima, A., Albanese, S., De Vivo, B., 2018. Measuring the change under Compositional Data Analysis (CoDA): insight on geochemical system dynamics. *J. Geochem. Explor.* 189, 100–108. <https://doi.org/10.1016/j.jgexplo.2017.05.006>.
- Cave, M.R., Johnson, C.C., Ander, E.L., Palumbo-Roe, B., 2012. Methodology for the determination of normal background contaminant concentrations in English soils. In: *British Geological Survey Commissioned Report, CR/12/003*, (41 pp.). <http://nora.nerc.ac.uk/19959/>.
- Cheng, Q., 1999. Spatial and scaling modelling for geochemical anomaly separation. *J. Geochem. Explor.* 65, 175–194.
- Cheng, Q., 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. *Ore Geol. Rev.* 32, 314–324.
- Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. *J. Geochem. Explor.* 51, 109–130.
- Cheng, Q., Xu, Y., Grunsky, E., 2000. Integrated spatial and spectrum method for geochemical anomaly separation. *Nat. Resour. Res.* 9, 43–51.
- Cheng, Q., Bonham-Carter, G.F., Raines, G.L., 2001. GeoDAS: a new GIS system for spatial analysis of geochemical data sets for mineral exploration and environmental assessment. In: *The 20th Intern. Geochem. Explor. Symposium (IGES)*. Santiago de Chile. Vol. 6/5–10/5. pp. 42–43.
- Cheng, Q., Xia, Q., Li, W., Zhang, S., Chen, Z., Zuo, R., Wang, W., 2010. Density/area power-law models for separating multi-scale anomalies of ore and toxic elements in stream sediments in Gejiu mineral district, Yunnan Province, China. *Biogeosciences* 7, 3019–3025.
- Chester, R., Stoner, J.H., 1973. Pb in particulates from the lower atmosphere of the eastern Atlantic. *Nature* 245, 27–28.
- Cicchella, D., De Vivo, B., Lima, A., 2005. Background and baseline concentration values of elements harmful to human health in the volcanic soils of the metropolitan and provincial area of Napoli (Italy). *Geochem. Explor. Environ. Anal.* 5, 29–40.
- De Vivo, B., Rolandi, G., Gans, P.B., Calvert, A., Bohrsen, W.A., Spera, F.J., Belkin, H.E., 2001. New constraints on the pyroclastic eruptive history of the Campanian volcanic Plain (Italy). *Mineral. Petrol.* 73, 47–65.
- De Vivo, B., Ayuso, R.A., Belkin, H.E., Fedele, L., Lima, A., Rolandi, G., Somma, R., Webster, J.D., 2003. Chemistry, fluid/melt inclusions and isotopic data of lavas, tephra and nodules from 25 ka to 1944 AD of the Mt. Somma-Vesuvius volcanic activity. In: *Mt. Somma-Vesuvius Geochemical Archive*. Dipartimento di Geofisica e Vulcanologia, Università di Napoli Federico II, Open File Report 1–2003, (143 pp).
- De Vivo, B., Petrosino, P., Lima, A., Rolandi, G., Belkin, H.E., 2010. Research progress in volcanology in Neapolitan area, Southern Italy: a review and alternative views. *Mineral. Petrol.* 99, 1–28. <https://doi.org/10.1007/s00710-009-0098-6>.
- De Vivo, B., Lima, A., Albanese, S., Cicchella, D., Rezza, C., Civitillo, D., Minolfi, G., Zuzolo, D., 2016. *Atlante geochimico-ambientale dei suoli della Campania* (Environmental Geochemical Atlas of Campania Soils). Aracne Editrice, Roma. pp. 364 (ISBN 978-88-548-9744-1. in Italian).
- Ducci, D., Tranfaglia, G., 2005. L'impatto dei cambiamenti climatici sulle risorse idriche sotteranee in Campania. *Boll. Ordine dei Geologi Della Campania* 1–4, 13–21 (in Italian).
- Duce, R.A., Hoffmann, G.L., Zoller, W.H., 1975. Atmospheric trace metals at remote northern and southern hemisphere sites: pollution or natural? *Science* 187, 59–61.
- Egli, M., Mirabella, A., Sartori, G., 2008. The role of climate and vegetation in weathering and clay mineral formation in late Quaternary soils of the Swiss and Italian Alps. *Geomorphology* 102, 307–324. <https://doi.org/10.1016/j.geomorph.2008.04.001>.
- Egozcue, J.J., Pawlowsky-Glahn, V., 2005. Groups of parts and their balances in compositional data analysis. *Math. Geol.* 37 (7), 795–828.
- Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barcelo-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. *Math. Geol.* 35 (3) (279–30).
- ESRI (Environmental Systems Research Institute), 2012. ArcGIS Desktop: Release 10. Redlands, CA.
- Filzmoser, P., Hron, K., 2008. Outlier detection for compositional data using robust methods. *Math. Geosci.* 40 (3), 233–248.
- Filzmoser, P., Hron, K., Reimann, C., 2009a. Principal component analysis for compositional data with outliers. *Environmetrics* 20 (6), 621–632.
- Filzmoser, P., Hron, K., Reimann, C., 2009b. Univariate statistical analysis of environmental (compositional) data - problems and possibilities. *Sci. Total Environ.* 407, 6100–6108.
- Gabriel, K.R., 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58 (3), 453.
- Goldberg, E.D., 1972. Baseline studies of pollutants in the marine environment and research recommendations. In: *The International Decade of Ocean Exploration (IDOEO) Baseline Conference*, 24–26 May 1972. New York.
- Hakanson, L., 1980. An ecological risk index for aquatic pollution control: a sedimentological approach. *Water Res.* 14, 975–1001.
- Han, J., Kamber, M., 2001. *Data Mining: Concepts and Techniques*. Morgan-Kaufmann Academic Press, San Francisco.
- Hron, K., Templ, M., Filzmoser, P., 2010. Imputation of missing values for compositional data using classical and robust methods. *Comput. Stat. Data Anal.* 54 (12), 3095–3107.
- ISTAT, 2016. Resident Population on 1st January, 2017. <https://www.istat.it/en/population-and-households>.
- Joron, J.L., Metrich, N., Rosi, M., Santocroce, R., Sbrana, A., 1987. Chemistry and petrography. In: Santacorce, R. (Ed.), *Somma-Vesuvius*. CNR Quad. Ric. Sci. Vol. 114. pp. 105–174.
- Lima, A., De Vivo, B., Cicchella, D., Cortini, M., Albanese, S., 2003a. Multifractal IDW interpolation and fractal filtering method in environmental studies: an application on regional stream sediments of Campania Region (Italy). *Appl. Geochem.* 18 (12), 1853–1865. [https://doi.org/10.1016/S08832927\(03\)00083-0](https://doi.org/10.1016/S08832927(03)00083-0).
- Lima, A., Danyushevsky, L.V., De Vivo, B., Fedele, L., 2003b. A model for the evolution of the Mt. Somma-Vesuvius magmatic system based on fluid and melt inclusion investigations. In: De Vivo, B., Bodnar, R.J. (Eds.), *Melt Inclusions in Volcanic Systems: Methods, Applications and Problems*. Series: Development in Volcanology Elsevier, Amsterdam 272 pp.
- Loska, K., Cebula, J., Pelczar, J., Wiechula, D., Kwapiński, J., 1997. Use of enrichment and contamination factors together with geoaccumulation indexes to evaluate the content of Cd, Cu, and Ni in the Rybnik water reservoir in Poland. *Water Air Soil Pollut.* 93, 347–365.
- Mandelbrot, B.B., 1983. *The Fractal Geometry of Nature*. Freeman, San Francisco, pp. 468.
- Maronna, R., Martin, R., Yohai, V., 2006. *Robust Statistics: Theory and Methods*. John Wiley, pp. 436 (ISBN: 978-0-470-01092-1).
- Martin, J.M., Whitfield, M., 1983. The significance of the river input of chemical elements to the ocean. In: Wong, C.S., Boyle, E., Bruland, K.W., Burton, J.D., Goldberg, E. (Eds.), *Trace Metals in Sea Water*. Plenum, New York, pp. 265–296.
- Minolfi, G., Albanese, S., Lima, A., Tarvainen, T., Fortelli, A., De Vivo, B., 2018. A regional approach to the environmental risk assessment - human health risk assessment: case study in the Campania region. *J. Geochem. Explor.* 184 (Part B), 400–416.
- Mondillo, N., Boni, M., Balassone, G., Rollinson, G., 2011. Karst bauxites in the Campania Apennines (southern Italy): a new approach. *Periodico di Mineralogia* 80, 407–432.
- Nziugheba, G., Smolders, E., 2008. Inputs of trace elements in agricultural soils via phosphate fertilizers in European countries. *Sci. Total Environ.* 390, 53–57.
- Olea, R.A., Raju, N.J., Egozcue, J.J., Pawlowsky-Glahn, V., Sing, S., 2018. Advancements in hydrochemistry mapping: Methods and application to groundwater arsenic and iron concentrations in Varanasi, Uttar Pradesh, India. *Stoch. Env. Res. Risk A.* 32 (1), 241–259.
- Otero, N., Tolosana-Delgado, R., Solera, A., Pawlowsky-Glahn, V., Canals, A., 2005. Relative vs. absolute statistical analysis of compositions: a comparative study of surface waters of a Mediterranean river. *Water Res.* 39, 1404–1414.
- Pawlowsky-Glahn, V., Buccianti, A., 2011. *Compositional Data Analysis: Theory and Applications*. John Wiley & Sons.
- Pawlowsky-Glahn, V., Egozcue, J.J., 2001. Geometric approach to statistical analysis on the simplex. *Stoch. Env. Res. Risk A.* 15 (5), 384–398.
- Peccherillo, A., 2005. Plio-quaternary volcanism in Italy. In: *Petrology, Geochemistry, Geodynamics*. Springer-Verlag, Berlin Heidelberg, pp. 365 (ISBN 978-3-540-29,092-6).
- Pearson, D.H., Cawse, P.A., Cambray, R.S., 1974. Chemical uniformity of airborne particulate material, and a maritime effect. *Nature* 251, 675–679.
- Pison, G., Rousseeuw, P.J., Filzmoser, P., Croux, C., 2003. Robust factor analysis. *J. Multivar. Anal.* 84, 145–172.
- Rahn, K.A., 1976. The chemical composition of the atmospheric aerosol. In: *Technical Report of the Graduate School of Oceanography*. University of Rhode Island, Kingston, R.I., USA.
- Reimann, C., De Caritat, P., 2005. Distinguishing between natural and anthropogenic sources for elements in the environment: regional geochemical surveys versus enrichment factors. *Sci. Total Environ.* 337, 91–107.
- Reimann, C., Filzmoser, P., Garrett, R., 2002. Factor analysis applied to regional geochemical data: problems and possibilities. *Appl. Geochem.* 17 (3), 185–206.
- Reimann, C., Garrett, R.G., Filzmoser, P., 2005. Background and threshold - critical comparison of methods of determination. *Sci. Total Environ.* 346, 1–16.
- Reimann, C., Filzmoser, P., Garrett, R.G., Dutter, R., 2008. Statistical data analysis

- explained. In: *Applied Environmental Statistics* with R. Wiley, Chichester, pp. 362 (ISBN: 978-0-470-98581-6).
- Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., O'Connor, P., GEMAS Team, 2014. Chemistry of Europe's agricultural soils — part A: methodology and interpretation of the GEMAS data set. In: *Geologisches Jahrbuch (Reihe B)*. Schweizerbarth, Hannover, pp. 528.
- Rolandi, G., Bellucci, F., Heizler, M.T., Belkin, H.E., De Vivo, B., 2003. Tectonic controls on genesis of ignimbrites from the Campanian Volcanic Zone, Southern Italy. In: De Vivo, B., Scandone, R. (Eds.), *Ignimbrites of the Campania Plain, Italy*. Mineral. Petrol. 79, pp. 3–31.
- Roosseeuw, P.J., Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41, 212–223.
- Saeedi, M., Li, L.Y., Salmanzadeh, M., 2012. Heavy metals and polycyclic aromatic hydrocarbons: pollution and ecological risk assessment in street dust of Tehran. *J. Hazard. Mater.* 227–228, 9–17.
- Salminen, R., Gregorauskiene, V., 2000. Considerations regarding the definition of a geochemical baseline of elements in the surficial materials in areas differing in basic geology. *Appl. Geochem.* 15, 647–653.
- Scheib, A.J., Birke, M., Dinelli, E., The Gemas Project Team, 2014. Geochemical evidence of aeolian deposits in European soils. *Boreas* 43 (1), 175–192. <https://doi.org/10.1111/bor.12029>.
- Sutherland, R.A., Tolosa, C.A., Tack, F.M.G., Verloo, M.G., 2000. Characterization of selected element concentration and enrichment ratios in background and anthropogenically impacted roadside areas. *Arch. Environ. Contam. Toxicol.* 38, 428–438.
- Tarvainen, T., Jarva, J., Johnson, C.C., Ottesen, R.T., 2011. Using geochemical baselines in the assessment of soil contamination in Finland. In: Demetriades, A., Locutura, J. (Eds.), *Mapping the Chemical Environment of Urban Areas*. Chichester, UK, John Wiley & Sons Ltd., pp. 223–231.
- Taylor, S.R., McLennan, S.M., 1995. The geochemical evolution of the continental crust. *Rev. Geophys.* 33, 241–265.
- Thiombane, M., Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Cavaliere, M., De Vivo, B., 2018a. Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model. *J. Geochem. Explor.* 189, 85–99.
- Thiombane, M., Martin-Fernandez, J.A., Albanese, S., Lima, A., Doherty, A., De Vivo, B., 2018b. Exploratory analysis of multi-element geochemical patterns in soil from the Sarno River Basin (Campania region, southern Italy) through compositional data analysis (CODA). *J. Geochem. Explor.* <https://doi.org/10.1016/j.gexplo.2018.03.010>.
- Tóth, G., Guicharnaud, R.-A., Tóth, B., Hermann, T., 2014. Phosphorus levels in croplands of the European Union with implications for P fertilizer use. *Eur. J. Agron.* 55, 42–52. (ISSN 1161-0301). <https://doi.org/10.1016/j.eja.2013.12.008>.
- U.S. Environmental Protection Agency (EPA), 2001. Guidance for characterizing background chemicals in soil at superfund sites, external review draft, Office of Emergency and Remedial Response, OSWER. 9285.7-41. In: Replaced by Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites, EPA 540-R-01-003, September 2002.
- Wedepohl, K.H., 1995. The composition of the continental crust. *Geochim. Cosmochim. Acta* 59, 1217–1232.
- Wu, S., Zhou, S., Li, X., 2011. Determining the anthropogenic contribution of heavy metal accumulations around a typical industrial town: Xushe, China. *J. Geochem. Explor.* 110, 92–97.
- Zuo, R., 2011. Decomposing of mixed pattern of arsenic using fractal model in Gangdese belt, Tibet, China. *Appl. Geochem.* 26, S271–S273.
- Zuo, R., Wang, J., 2016. Fractal/multifractal modelling of geochemical data: a review. *J. Geochem. Explor.* 164, 33–41.
- Zuo, R., Wang, J., Chen, G., Yang, M., 2015. Identification of weak anomalies: a multifractal perspective. *J. Geochem. Explor.* 148, 12–24.
- Zuzolo, D., Cicchella, D., Catani, V., Giaccio, L., Guagliardi, I., Esposito, L., De Vivo, B., 2017. Assessment of potentially harmful elements pollution in the Calore River basin (southern Italy). *Environ. Geochem. Health* 39, 531–548. <https://doi.org/10.1007/s10653-016-9832-2>.

Paper 2

Soil Compositional Contamination Index: a new approach to reveal and quantify contamination patterns of potentially toxic elements in Campania Region (Italy)

Attila Petrik, **Matar Thiombane**, Annamaria Lima, Stefano Albanese, Benedetto De Vivo

Journal of applied geochemistry, Volume 96, September 2018, Pages 264-276



Soil contamination compositional index: A new approach to quantify contamination demonstrated by assessing compositional source patterns of potentially toxic elements in the Campania Region (Italy)

Attila Petrik^{a,*}, Matar Thiombane^a, Annamaria Lima^a, Stefano Albanese^a, Jamie T. Buscher^{b,c}, Benedetto De Vivo^{d,e}

^a Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse, Università degli Studi di Napoli "Federico II", Complesso Universitario Monte S. Angelo, Via Cintia snc, 80126, Naples, Italy

^b Andean Geothermal Center of Excellence (CEGA), Universidad de Chile, Plaza Ercilla 803, Santiago, Chile

^c Department of Geology, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Plaza Ercilla 803, Santiago, Chile

^d Pegaso University, Piazza Trieste e Trento 48, 80132 Napoli, Italy

^e Benecon Scarl, Dipartimento Ambiente e Territorio, Via S. Maria di Costantinopoli 104, 80138 Napoli, Italy

ARTICLE INFO

Editorial handling by C. Reimann

Keywords:

Campania region (Italy)
Potentially toxic elements
Spatial abundance
Source discrimination
Soil contamination compositional index

ABSTRACT

Potentially toxic elements (PTEs) are a major worldwide threat to the environment due to the constant global increase in industrial activity and urbanisation. Several studies have provided detailed maps and a better understanding of the spatial distribution patterns of PTEs in different matrices, but the majority of these studies have simply neglected the compositional nature of geochemical data. The aims of this study are to reveal the compositional behaviour and relative structure of 15 PTEs (subcomposition) in Campania, one of the most contaminated regions in Italy, and to quantify the spatial abundance and identify the possible origins of these PTEs. Robust compositional biplots were used to understand the natural grouping and origin of the PTEs. Ratios of specific subcompositions (balances) of PTEs were calculated to map the spatial patterns and identify the spatial variability of the PTEs. This study presents the preliminary steps needed to quantify and analyse the relative difference in the spatial abundance of PTEs by applying a compositional abundance index. In addition, a new soil contamination compositional index (SCCI) was elaborated to quantify topsoil contamination by the 15 PTEs and related subgroups following the compositional structure of the geochemical data.

The elevated spatial abundance of the 15 PTEs is related to highly urbanised (Naples and Salerno), highly industrialised (Solofra) and intensely cultivated areas (Sarno River Basin), where the high dominance of elements from the anthropogenic subgroup (Pb, Sb, Sn and Zn) and high SCCI values suggest that contamination is from anthropogenic sources. The high spatial dominance of elements from the volcanic rock subgroup (As, Be, Se, Tl and V) in these same areas is likely related to geogenic sources, including alkalic pyroclastic rocks. Although the high spatial abundance of Group B elements (Cd, Cr, Co and Ni) is related to Terra Rossa soils and shaley facies of siliciclastic rocks of the southern Apennines, these same elements can also reach high abundances and reflect contamination (i.e. high SCCI values) from urbanised and industrialised areas due to e.g., tanneries and alloy production.

Other high spatial abundances of the 15 PTEs with little or no contamination (i.e. very low SCCI values) can be related to nearby carbonate massifs, where a mixture of geogenic factors including weathering, advanced pedogenic processes, adsorption and co-precipitation with Fe-/Mn-oxyhydroxides and the presence of pyroclastic material might all be responsible for an increase in abundance.

The lowest spatial dominance of the 15 PTEs occurs in the northeastern and southwestern siliciclastic zones of the Campania Region, where there is a low level of urbanisation and industrialisation and therefore contamination from any source can be excluded.

* Corresponding author. Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse, Università degli Studi di Napoli "Federico II", Complesso Universitario Monte S. Angelo, Via Cintia snc, 80126, Naples. Italy.

E-mail address: tectonic@caesar.elte.hu (A. Petrik).

<https://doi.org/10.1016/j.apgeochem.2018.07.014>

Received 28 March 2018; Received in revised form 19 July 2018; Accepted 21 July 2018

Available online 23 July 2018

0883-2927/ © 2018 Elsevier Ltd. All rights reserved.

1. Introduction

One of the main objectives of environmental geochemistry research is to constrain the spatial distribution patterns of various organic and inorganic elements and compounds to determine whether sources are geogenic and/or anthropogenic. Geochemical mapping and the delimitation of mineralisation or contamination/pollution areas has always been the focus of exploration and environmental geochemistry. Several graphical techniques including proportional dots (Björklund and Gustavsson, 1987), statistical methods like Mean + 2 Standard Deviation (Hawkes and Webb, 1962; Reimann and Garrett, 2005) and Median + 2 Median Absolute Deviation (Reimann et al., 2005), cumulative probability plots (Tennant and White, 1959; Lepeltier, 1969; Sinclair, 1974), upper-fence criteria in Tukey's boxplots (Tukey, 1977) and stated percentiles (e.g., 95th, 98th) in box-and-whiskers plots (Kürzl, 1988) have all been used to define geochemical threshold values and outliers. The separation of elemental anomalies from baseline and background values using various fractal methods such as the Concentration-Area (C-A) or Spectrum-Area (S-A) techniques and local singularity analysis have been successfully applied (Cheng, 1999; Cheng et al., 1994, 2000; Lima et al., 2003a, 2005; Albanese et al., 2007, 2015; Zuo et al., 2015; Zuo and Wang, 2016; Parsa et al., 2017; Minolfi et al., 2018; Rezza et al., 2018).

Potentially Toxic Elements (PTEs) have always been given special emphasis because the accumulation of these elements in different matrices can cause soil and land degradation that can then be transferred to the human body from dermal contact, inhalation or ingestion through the food chain and drinking water (Lim et al., 2008; Ji et al., 2008; Varrica et al., 2014). PTEs are a major environmental concern because concentrations are constantly rising due to accelerated population growth and therefore increased urbanisation and industrialisation that generate a wide range of anthropogenic contamination sources (Albanese et al., 2010; Guillén et al., 2011; Wang et al., 2012; Wu et al., 2015). PTE compounds are generally non-biodegradable with long biological half-lives that tend to accumulate in soils by adsorption to clay minerals and organic matter (Kabata-Pendias, 2011). However, PTE bioavailability is influenced by different physicochemical processes (e.g., pH, Eh) and physiological adaptations (Yang et al., 2004; Skordas and Kelepertsis, 2005; Barkouch et al., 2007; Kabata-Pendias, 2011; Zhao et al., 2013, 2014).

Several single and integrated contamination indices have been used, such as the Enrichment factors (Chester and Stoner, 1973), Geoaccumulation Index (Muller, 1969) and Single Pollution Index (Hakanson,

1980; Muller, 1981), to quantify the contamination status of different environmental media. Indices using background/baseline values (e.g., Single Pollution Index) for reference are straightforward, but are not scale-invariant meaning that the change in units of the concentrations modifies the results of the analysis (Aitchison and Egozcue, 2005; Pawłowsky-Glahn et al., 2015). The critical reviews of element ratio variations and Enrichment factors (EFs) were made by Reimann and de Caritat (2000, 2005) claiming that their values vary and are dependent on the different parent rock materials and chosen reference media as well as reference elements. In addition, these indices do not take into account the different biogeochemical processes, the natural fractionation of elements or differential solubility of minerals which may have remarkable impact on elemental enrichment/contamination (Reimann and de Caritat, 2000, 2005). Sucharovà et al. (2012) emphasised that high EFs or top-/bottom soil ratios reflect the geochemical decoupling of the lithosphere from the biosphere rather than contamination. They concluded that regional distribution of the raw data from multi-medium provides a better indication of the extent and impact of pollution than ratios calculated based on misconceptions.

Fabian et al. (2017) introduced a new method for detecting and quantifying diffuse contamination based on the analysis of cumulative distribution functions. This method revealed local contamination sources and efficiently monitored diffuse contamination at the continental to regional scale.

However, geochemical data are inherently compositional data; therefore indices measuring the contamination/pollution should be considered the main principles of compositional data analysis.

This study aims to better understand the compositional behaviour and relative structure of a subcomposition of 15 elements, which are listed as PTEs for the Campania Region (D. Lgs. 152/2006), and to quantify the spatial abundance and identify the possible origins of PTEs.

We put special emphasis on the ratios of specific subcompositions (balances) of PTEs by mapping the spatial patterns and making profiles to reveal the spatial variability. This study presents the preliminary steps to analyse the spatial abundance of PTEs by applying a compositional abundance index. A new soil contamination compositional index was elaborated to quantify the topsoil contamination from the 15 PTEs and associated subgroups following the compositional structure of geochemical data.

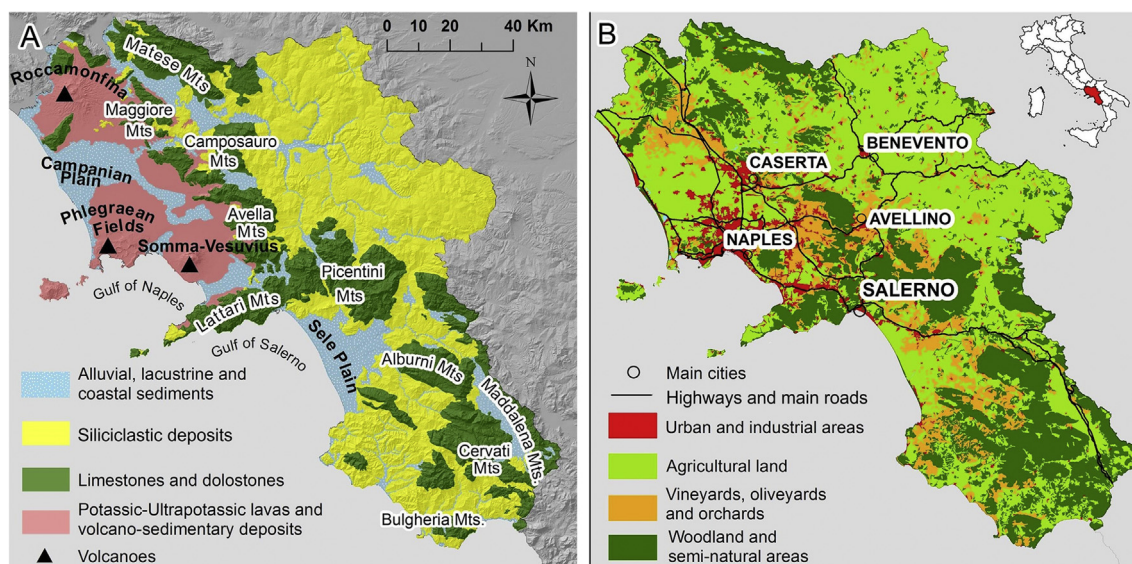


Fig. 1. Simplified geological (A) and landuse (B) maps of the Campania Region.

2. Study area

2.1. Geology

The Campania Region is located in southwestern Italy and occupies an area of about 13,600 km² (Fig. 1) that can be divided into three different morphological and geological areas. The high mountainous area of the region is part of the NW-SE trending southern Apennines that evolved during the Cretaceous - Tertiary Alpine orogeny (e.g., Bonardi et al., 2009; Vitale and Giarcia, 2013). This part of the southern Apennines is mostly composed of sedimentary and volcanic rocks spanning in age from the Triassic to the Present (e.g., Bonardi et al., 2009) (Fig. 1A). The highest mountains in the region (1800–2000 m) are carbonatic massifs (e.g., Mt. Matese, Mt. Picentini) that consist of Mesozoic limestone and dolostone (Fig. 1A).

The southern Apennines are bound by hilly topography dominated by siliciclastic rocks (e.g., sandstone, siltstone, claystone, flysch) that were deposited in different tectonic settings ranging from deep to shallow water and terrigenous environments between the Mesozoic and Tertiary (e.g., Peccerillo, 2005; Buccianti et al., 2015) (Fig. 1A).

The western part of the region is characterised by relatively flat low-lying areas that include alluvial and coastal plains (e.g., Volturno, Sele and Campanian) (Fig. 1A).

These plains are structural depressions filled with Quaternary alluvial, coastal and volcanoclastic sediments that can reach a thickness of 5 km in some locations (Ippolito et al., 1973). Volcanic rocks consist of potassic and ultrapotassic lavas and pyroclastic deposits that range in age from 0.6 to 0.1 Ma (Giannetti and Luhr, 1983; Peccerillo, 2005; De Vivo et al., 2010). The majority of the pyroclastic material was generated by large ignimbrite events of Mt. Somma-Vesuvius and the Phlegraean Fields (e.g., Campanian Ignimbrite, 39 ka ago) (Rolandi et al., 1998, 2003; De Vivo et al., 2001; Torrente and Milia, 2013). Alkaline magmatism characterised by a high K content (Ayuso et al., 1998; De Vivo et al., 2010) is part of the Roman Co-magmatic Province (Washington, 1906).

2.2. Landuse

Campania is one of the most populated regions in Italy with more than 5.8 million inhabitants (ISTAT, 2016). One of the most important sectors of the economy is agriculture, which covers more than 50% of the total land area in the region (Fig. 1B). The agricultural lands are mainly concentrated in the fertile alluvial and coastal plains in the west (e.g., Campanian Plain), where extensive use of fertilisers and pesticides has resulted in severe contamination by organic and inorganic compounds, especially in the Sarno Basin (Cicchella et al., 2005; Albanese et al., 2007; Qu et al., 2017). In the terraced lands surrounding the southern Apennines, vineyards and olive tree orchards are dominant (Fig. 1B).

Industrial areas are scattered throughout the region but are mostly concentrated around large urbanised regions like Naples, Salerno and Avellino (Fig. 1B). The majority of the industrial processes are related to agricultural products, clothing materials and tanning (Albanese et al., 2007; De Vivo et al., 2016).

Transport networks are the most dense in the western urbanised areas around Naples and Salerno where vehicle emissions cause severe air pollution, especially in the summer when traffic jams occur frequently (Fig. 1B).

3. Data and methods

3.1. Sampling

Topsoil samples have been collected across the Campania Region with a nominal density of 1 sample/4 km² (16 km² sampling grid in suburban areas and 3.8 km² in urban areas). At each sampling site,

1.5 kg of soil was collected from a depth between 0 and 20 cm below the ground surface after removal of the vegetation cover. The sampling procedure and sample preparation followed the Geochemical Mapping of Agricultural and Grazing Land Soils (GEMAS) sampling protocol described by Reimann et al. (2014).

The samples were analysed by the ultratrace aqua regia extraction method (Bureau Veritas Minerals, 2017), using a combination of inductively coupled plasma atomic emission spectrometry (ICP-AES) and inductively coupled plasma mass spectrometry (ICP-MS) analyses to determine the “pseudototal” concentration of 53 elements. The “pseudototal” corresponds to the part concentration of elements that are extractable using an aqua regia digestion. It estimates the maximum amount of metals that could hypothetically be mobilized and transported in the environment from geogenic and anthropogenic sources (Vercoetere et al., 1995; Adamo and Zampella, 2008). A more detailed description of the sampling and analytical methods is given by Buccianti et al. (2015).

Of the 53 elements analysed, we only focused on the complete cases (2333 samples) of the 15 PTEs (subcomposition) to better understand and reveal the relative structure and spatial abundance. For three of the 15 elements considered some samples showed concentrations below the respective detection limit: Cr: 0.04% (DL: 0.5 mg/kg), Sb: 0.22% (DL: 0.02 mg/kg) and Se: 5.11% (DL: 0.1 mg/kg) of all samples. Concentrations below the detection limit were set to 65% of the detection limit for these elements as suggested by Palarea-Albaladejo and Martín-Fernández (2013).

3.2. Methods

3.2.1. Statistical analysis

A wide range of summary statistics [lower and upper quartiles (Q1, Q3), maximum, minimum, mean, robust coefficients of variation (CVR), median absolute deviation (MAD), kurtosis and skewness] were applied in this study using log-transformed data that was then back-transformed to describe the central tendency and variability of the investigated PTEs. Although the log-ratio transformation of data is more relevant in compositional data analysis, the summary statistics output expressed in the raw concentrations of single elements is also meaningful and more easily interpretable.

Tukey boxplots were also generated with logscaling to visualize the distributions and distinguish the univariate outliers of each of the 15 PTEs. The intervention limits of the 15 PTEs for residential and industrial areas established by Italian law (D. Lgs. 152/2006) were also displayed. To serve as an additional reference, the median values of the 15 PTEs from European agricultural soils (Reimann et al., 2014) were also included and the ratio between the Campanian topsoil medians and European counterparts was calculated.

3.2.2. Compositional multivariate analysis

Geochemical data are typically compositional in nature with positive vectors whose elements (parts) describe quantitatively relative contributions to a whole (Aitchison, 1986; Pawlowsky-Glahn and Buccianti, 2011; Pawlowsky-Glahn et al., 2015). At each sampling or interpolation point, the concentration data provides information on the relative weight of one particular element with respect to the total (Tolosana-Delgado and Van den Boogaart, 2013). One challenge with geochemical data is the problem of closure as the concentrations sum to a constant (e.g., as a percentage or in ppm), with an increase in one element concentration causing a decrease in other element concentrations (Chayes, 1960). The closure effect can be clearly observed in data of major elements that sometimes have concentrations that are orders of magnitude higher than those of trace elements (McKinley et al., 2016). In addition, the absolute concentration values are not scale-invariant, which primarily affects components with lower concentrations like trace elements.

Compositional data can be analysed in the simplex space, which

consists of D-parts that add up to make a constant sum. As an alternative approach, Aitchison (1986) elaborated several log-ratio transformations to open the data structure and conduct analyses in real Euclidean space. For example, pairwise, additive (alr) and centred log-ratio (clr) transformations of raw data can be analysed using classical statistical tools because these transformed data are scale-invariant and therefore devoid of closure effects. Nevertheless, limitations do exist for these transformations. For example, the ratio of the alr transformation is strongly dependent on the common denominator element and they are not coordinates with respect to orthonormal basis (McKinley et al., 2016). The clr transformation resolves the problem of the common denominator, but the geometric mean of the composition is affected by dilution or enrichment of some parts and the centred log-ratios are not coordinates with respect to orthonormal basis (Egozcue and Pawłowsky-Glahn, 2006).

Egozcue et al. (2003) introduced the isometric log-ratio (ilr) transformation to generate coordinates with an orthonormal basis, which cannot be affected by intrinsic distortion of scatterplot matrices and are subcompositionally coherent. One way to get ilr coordinates is to use balances, which are normalised log-ratios of the geometric means of two groups of elements (Egozcue et al., 2003). Balances can be obtained by choosing a sequential binary partition of an association of elements. Balances can be calculated using the following equation:

$$z_i = \sqrt{\frac{r_i s_i}{r_i + s_i}} \ln \frac{(\prod_{+} X_j)^{\frac{1}{r_i}}}{(\prod_{-} X_k)^{\frac{1}{s_i}}} \quad \text{for } i=1, \dots, D-1, \quad (1)$$

where the products \prod_{+} and \prod_{-} only include parts coded with a + and - sign, and r_i and s_i are the numbers of positive and negative signs (parts) in the i -th order partition, respectively (Egozcue and Pawłowsky-Glahn, 2005). The two separated groups of elements (parts) were used in the calculation of the given balance. The empty entries are parts that were not involved in the partition in the i -th order. Balances were calculated using Eq. (1) with the CoDAPack 2.02.21 software (Comas-Cuñí and Thió-Henestrosa, 2011).

Balances can be interpreted by considering three scenarios: (1) a positive balance when parts (variables) in the numerator have higher dominance with respect to parts in the denominator; (2) a negative balance when parts involved in the numerator have lower dominance than those in the denominator; (3) a nearly zero balance when the dominance of the two groups of parts is similar.

Groups of elements in these balances were chosen using unsupervised approaches like robust-biplot analysis. Biplots display both samples and variables of a data matrix in terms of the resulting scores and loadings (Gabriel, 1971). Aitchison and Greenacre (2002) presented a clr-biplot methodology for compositional data using singular-value decomposition to obtain a two-rank approximation of a centred log-ratio (clr) transformed data matrix. They introduced the term relative variation biplot, which represents the variation in all of the component ratios and respects the principle of subcompositional coherence of compositional data. Subcompositional coherence states that results obtained from the analysis of a composition cannot be contradictory to those obtained from the analysis of a subcomposition (Aitchison and Greenacre, 2002). The scores on a biplot represent the structure of the compositional data in Euclidean space based on the covariance matrix and additionally display the association structure of the dataset. The loadings (rays) represent the corresponding clr variables drawn from the centre of the plot, with the length being proportional to the amount of their explained variance (communality). The interpretation of the plot depends on the loadings (ray lengths) structures and more specifically on the approximate links between rays and samples and the distances between vertices and the relative directions (Otero et al., 2005).

However, outliers influence the results of principal components and can distort the covariance structure, so a robust version of the biplot was developed by Filzmoser and Hron (2008, 2009). In the robust

version, scores and loadings are calculated in orthonormal (ilr) coordinates to avoid singularity and the results then back-transformed to clr-space using the linear relation between clr and orthonormal coordinates (Egozcue et al., 2003). In the robust version, the minimum covariance determinant (MCD) is used to obtain the location and scatter estimators through a subset of at least h observations with the smallest determinant of the sample covariance matrices (Rousseeuw and Van Driessen, 1999). The number choice for h determines the robustness and efficiency of the estimators such that taking half of all of the data as a subset will yield a maximum resistance to outliers but a poorer efficiency. As a compromise, the value of h can be approximated as $\frac{3}{4}$ of all of the data (Filzmoser et al., 2009). The robust biplot was created by considering not only the 15 PTEs but also 9 major elements (Al, Ca, Fe, K, Mg, Mn, Na, P and Ti). The major elements were included to better interpret and delineate the subgroups of PTEs and distinguish geogenic factors from anthropogenic ones. However, for the construction of balances, the focus was only on the relative structure of the 15 PTEs (subcomposition).

Based on the robust biplot analysis and taking into account the local geological setting and geochemical properties of elements, a sequential binary partition was performed using the 15 PTEs subcomposition. We were particularly interested in the main contamination sources and spatial patterns of the PTEs, which can be characterised using 4 specific balances (Z1, Z6–Z8). Each of these had elements belonging to distinct groups on the robust biplot, reflecting different geogenic and anthropogenic processes.

Interpolated maps were generated from these balances using the multifractal inverse distance weighted (MIDW) method (Cheng et al., 1994) with GeoDAS software. MIDW is one of the most widely used methods for interpolating geochemical data because it preserves high frequency information and retains local variability by taking into account both spatial association and local singularity (Cheng, 1999; Cheng et al., 1994; Lima et al., 2003a, 2005). For the interpolation, we used a 5 km search radius and a minimum of 4 data points, with the map resolution set to 1 km based on the break point of the nearest distance cumulative plot of the sampling sites. Considering that the balance values can be negative and therefore not “log transformable”, a min-max normalisation was applied by scaling the original data within a specified range of values (e.g., ranging from 1 to 100). Min-max normalisation is a linear transformation applied to the original data that does not affect the geometrical structure (Han and Kamber, 2001). After interpolation, the normalised values were back-transformed to visualize the original range of balances.

In addition, the concentration–area (C–A) fractal plot (Cheng et al., 1994) was used to classify the interpolated balance maps (Z1, Z6–Z8) and capture the different spatial patterns. The C–A method is widely used to recognise different spatial patterns of elemental concentrations by taking into account not only the frequency distribution of concentration values but also the geometrical and spatial properties of identified features (Cheng et al., 1994). Profiles were also made for the interpolated balance maps in different directions to capture spatial variability and define outliers of specific groups of elements.

3.2.3. Compositional abundance maps and the soil contamination compositional index

Soil contamination by PTEs represents one of the foremost environmental issues in the world today. High concentrations of PTEs and associated compounds put human beings at risk through dermal contact, inhalation and ingestion (Lim et al., 2008; Ji et al., 2008; Varrica et al., 2014). A compositional index should be invariant from a change in measurement units or multiplication by a constant (Egozcue, 2009). Compositional data can be properly analysed using log-contrasts as a balance between certain groups of elements, as is the case for pollution indices that consider the ratio of the geometric mean of pollutant concentrations with that of non-pollutant concentrations (Jarauta-Bragulat et al., 2015). Jarauta-Bragulat et al. (2015) also introduced a

new air quality index (AQI) that was a good approximation of a log-contrast as a balance.

AQI was calculated as $AQI(x) = k^* g_{n-1}(x)$, where k^* is a constant depending on the units of pollutant concentrations and $g_{n-1}(x)$ is the geometric mean of air pollutant concentrations. In order to get a normalised index, they used an arbitrarily large geometric mean value ($\max g_{n-1}(x)$) as a denominator which correspond to an almost incredible pollution of the air and modified the equation to be the following:

$$AQI^* = (k^* / \max g_{n-1}(x)) * g_{n-1}(x) \quad (2)$$

In this study, the same equation as Eq. (2) was tested on the 15 PTEs but we used the maximum geometric mean of the PTEs as a denominator and changed the name to Compositional Abundance Index (CAI). Compositional abundance indices were also calculated for specific subgroups that were categorized as the following: Volcanic (As, Be, Tl, V and Se), Terra Rossa soil and Siliciclastic (Cr, Ni, Co and Cd) and Anthropogenic (Pb, Sb, Sn and Zn) Compositional Abundance Indices. The calculation was applied to subgroups based on a sequential binary partition table using Eq. (2). We used 100 as the k^* constant in Eq. (2) to get abundance ranges between 0 and 100. Although the CAI is compositionally meaningful and easily interpretable, the maximum geometric mean of elemental concentrations in the denominator makes it a one-off estimator for a survey. This was the motivation for elaborating a new Soil Contamination Compositional Index (SCCI) for both the total and subgroups of the 15 PTEs. The SCCI is stable and retains the same value for older observations even if the geochemical sampling is expanded. The SCCI was calculated using the following equation:

$$SCCI = g_{n-1}(x) / g_{n-1}(x_{bas}) \quad (3)$$

where the geometric mean of the 15 PTEs concentrations is $g_{n-1}(x)$ and the geometric mean of the baseline values of these elements is $g_{n-1}(x_{bas})$. We used the baseline values of each of the 15 PTEs published by De Vivo et al. (2016) for the entire Campania Region. These baseline values for each element were obtained by the Spectrum-Area (S-A) fractal method developed by Cheng et al. (1994, 2000). Although baseline values can be decomposed into anthropogenic and geogenic parts, the larger the SCCI is, the larger the contamination of topsoils may be. SCCI values close to 1 indicate little or no contamination by the investigated PTEs. The SCCI was also calculated for the same specific subgroups (Volcanic, Terra Rossa soil and Siliciclastic, Anthropogenic) using the respective baseline values to reveal contamination source patterns. The SCCI ratio values were interpolated using the MIDW method and classified by C-A fractal plots.

4. Results and discussion

4.1. Univariate statistical analysis

Table 1 displays the summary statistics of the 15 PTEs and the percentages below the detection limit for Cr, Sb and Se. In terms of variability, Cu has the highest robust coefficient of variation (CVR) (56.45%), followed by Se (50%) and Hg (46.15%). The lowest CVR is represented by Ni (26.45%), Co (26.67%) and As (26.89%). The distribution of variables is mainly right-skewed and not symmetric based on positive skewness and kurtosis values (Table 1).

The Tukey boxplots also reveal the asymmetry in the distribution of variables and the dominance of high outliers on a logscale (Fig. 2). The medians of variables of the Campanian topsoil data exceed the European counterparts (GEMAS, Reimann et al., 2014) for all of the elements except for Cr. The elements with the highest ratios of medians are Tl ($r = 11.6$), Be and Sn ($r = 4.58$) (Fig. 2), while the elements with the lowest ratios of medians are Cr ($r = 0.8$) and Ni ($r = 1.03$). All of the 15 PTEs have concentration values above the intervention limits for residential areas and a few samples also exceed the industrial limits set by

Italian law (D. Lgs. 152/2006) (Fig. 2). The analyses for Cu, Zn, Pb, As, Be, Sn and Tl are the most concerning because the vast majority of high outliers are above the residential intervention limits, indicating that there are potential risks for human beings (Fig. 2). It should be emphasised that Sn becomes toxic only if combined with organics to form organo-Sn compounds (Kabata-Pendias, 2011), though Italian Law reports it as a PTE. The average upper-crustal concentrations of the 15 PTEs (Wedepohl, 1995) are generally lower than the medians of the respective elements in the Campania Region with the exception of Cr, Ni and Co (Fig. 2).

4.2. Compositional multivariate analysis

4.2.1. Robust biplot analysis

The first two axes of the compositional robust biplot explain 60.58% of the total variability of the 15 PTEs and 9 major elements (Fig. 3). The biplot reveals several distinct groups, but special emphasis was put on the PTEs.

Group A consists of the clr (centred log-ratio) coefficients of Hg, Pb, Sb and Sn, with rays pointing in the same direction with the highest loading of the clr-Hg (Fig. 3). This group is represented by clr-coefficients that are likely related to anthropogenic activity including heavy traffic, fossil fuel combustion, illegal waste disposal and extensive agricultural activity. The anthropogenic origin of these elements was also revealed in stream sediment data by Albanese et al. (2007) and in other local-scale studies by Cicchella et al. (2005, 2016) and Petrik et al. (2018a).

The clr-Hg might also be influenced by a combination of processes because the vertex deviates from the rest of Group A and reveals a high clr-variability (vector length) (Fig. 3). High Hg concentration has been attributed to anthropogenic (e.g., traffic, fossil fuel combustion) and hydrothermal activity (Cicchella et al., 2005; De Vivo et al., 2016).

Group B is represented by the clr-coefficients of 4 PTEs (Ni, Cr, Co and Cd) and 2 major elements (Fe and Mn) (Fig. 3A). The rays are pointing in the same direction with the longest vector length of Cr and Ni (Fig. 3A). Cd has the lowest clr-variability of the 4 PTEs. The association of these 4 PTEs with Fe and Mn may be related to phyllosilicates and Fe-/Mn-sulphides from black shaley facies of siliciclastic rocks (e.g., Crete Nere Fm.). However, the biplot projected on the PC1-PC3 axes reveals that these 4 PTEs are also accompanied by the major elements of Ca and Mg (Fig. 3B). This association reflects different geogenic processes that may be indicative of elements sourced from Terra Rossa soils. Terra Rossa soils (e.g., Haplic Luvisols) were mapped (Di Gennaro et al., 2002) throughout the NW-SE trending carbonate massifs of the southern Apennines in the Campania Region where some bauxite mineralisation was also found (Mondillo et al., 2011; Boni et al., 2013). In fact, the 4 PTEs of Group B seem to be associated not only with shaley facies of the siliciclastic rocks but also with the Terra Rossa soils.

Group C forms a separate association that includes the clr-coefficients of As, Be, Tl, V and Se and the major elements of Al and Ti. The rays are scattered but all point in the same direction with the longest vector length of Ti (Fig. 3A). These clr-coefficients may be related to the presence of volcanoclastic rocks and hydrothermal/volcanic activity (Tarzia et al., 2002; Lima et al., 2003b; De Vivo et al., 2016). However, the relative abundance and spatial pattern analysis of As and Be revealed that these elements may be at least partially associated with advanced pedogenic processes (e.g., mature Andisols found with a Chemical Index of Alteration of 70–80) in the Campania Region (Petrik et al., 2018b; c). De Vivo et al. (2016) related some of these elements (e.g., Al, As, Be and Tl) to the oldest volcanic products of the Campanian eruption events.

Group D contains only the clr of Cu and the major element of P, with vertices separated from other associations that point in a different direction. Studies of the stream sediment and topsoil from the Campania Region and Sarno River Basin demonstrated that these elements are related to agricultural activity using sewage sludge and phosphate

Table 1

Summary statistics of the 15 investigated PTEs based on log-transformed data that was then back-transformed.

PTEs	Min	Q1	Mean	Mean-log	Median	Q3	Max	CVR (%)	MAD	Skewness	Kurtosis	< DL (%)
Cu	2.51	31.89	97.30	63.45	58.57	124.19	2394.33	56.45	33.25	0.19	−0.16	
Pb	3.12	32.07	67.09	49.09	51.50	73.34	1305.91	39.16	20.23	0.12	1.32	
Zn	11.40	69.10	117.25	96.36	89.60	126.75	3210.60	28.76	25.70	0.80	2.77	
Ni	0.50	12.10	18.39	15.22	15.50	20.90	98.90	26.45	4.10	−0.48	1.35	
Co	0.80	7.60	10.81	9.83	10.50	13.30	79.00	26.67	2.80	−0.49	1.31	
As	0.60	8.10	11.97	10.51	11.90	14.80	111.50	26.89	3.20	−0.89	1.99	
V	5.00	43.00	63.71	57.52	59.00	84.00	224.00	32.20	19.00	−0.65	0.65	
Cr	0.325	11.10	20.20	15.73	16.00	23.10	808.40	35.00	5.60	−0.16	2.99	0.04
Cd	0.02	0.23	0.49	0.37	0.37	0.56	11.06	43.24	0.16	0.27	1.24	
Tl	0.05	0.74	1.37	1.06	1.40	1.95	3.62	42.86	0.60	−1.24	0.87	
Sb	0.013	0.43	0.97	0.65	0.65	0.98	40.79	40.00	0.27	0.14	3.43	0.22
Hg	0.003	0.033	0.094	0.058	0.052	0.098	6.775	46.15	0.025	0.50	0.97	
Se	0.065	0.20	0.43	0.33	0.40	0.60	2.40	50.00	0.20	−1.04	0.46	5.11
Sn	0.20	1.90	3.97	2.98	3.30	4.70	125.60	42.42	1.40	−0.27	0.63	
Be	0.20	2.70	4.43	3.71	4.50	5.90	16.90	33.33	1.50	−1.21	1.29	

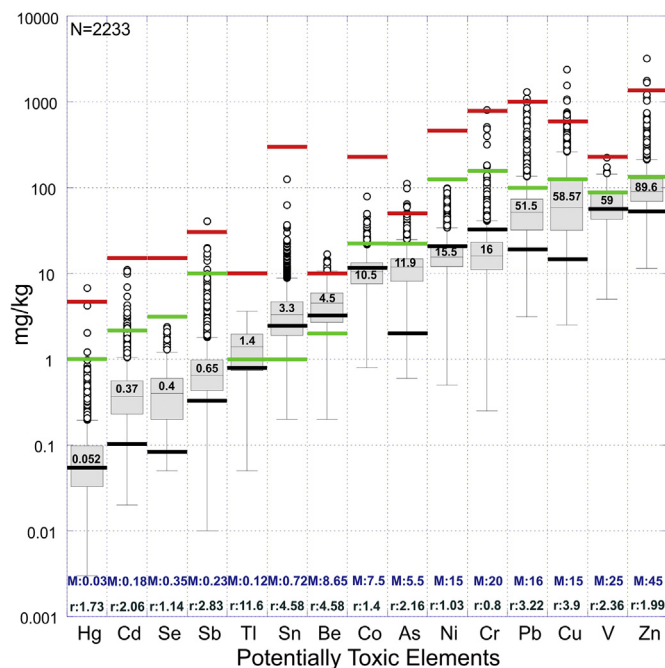


Fig. 2. Box-plots of the 15 PTEs. Green and red lines show the intervention limits for residential and industrial areas, respectively (D. Lgs. 152/2006). The thick black lines mark the average upper crustal concentration of the given element (Wedepohl, 1995). The numbers in the boxes are the medians of the respective elements. M: medians of respective elements from the Geochemical Mapping of European Agricultural Soil project (Reimann et al., 2014). r: ratios between the medians of the Campanian PTEs and the European counterparts. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

fertilizers (Albanese et al., 2007; Adamo et al., 2014; Cicchella et al., 2016).

Finally, the last association includes the clr-coefficients of Na and K that are not reported as PTEs by Italian law (D. Lgs. 152/2006) and have high abundances related to the youngest volcanic rocks of Mt. Somma-Vesuvius and the Phlegraean Fields (De Vivo et al., 2016).

4.2.2. Balance maps (Z1, Z6–Z8)

Based on the robust biplot analysis, a sequential binary partition was performed that involved the 15 PTEs subcomposition (Table 2). There was a particular interest in groups of elements that represented different geogenic and/or anthropogenic processes. For this reason, 4 balance maps (Z1, Z6–Z8) were further analysed to reveal spatial patterns and contamination sources (Table 2).

4.2.2.1. Z1 (ilr-1) balance map. The Z1 balance map illustrates that the abundance of the volcanic rock elements (Group C: As, Be, Tl, V and Se) with respect to all of the other PTEs (Fig. 4A, Table 2). The vast majority of Z1 balance values are below 0, meaning that the dominance of volcanic rock elements is lower than other PTEs in the study area (Fig. 4A). The highest spatial abundance of volcanic rock elements (−0.04–0.88) is related to large volcanic complexes (e.g., Roccamonfina, Phlegraean Fields and Mt. Somma-Vesuvius) and carbonate massifs (e.g., Mt. Picentini and Mt. Cervati) overlain by pyroclastic deposits (Fig. 4A, C-Profiles 1 and 2).

The lowest spatial abundance (−3.28–1.87) was from topsoils found over siliciclastic deposits in the northeastern and southwestern parts of the Campania Region (Fig. 4A). Low spatial abundances of volcanic rock elements were also revealed along large rivers like the Volturno, Sele and Calore, where other non-volcanic elements have accumulated by river transport thereby disturbing the spatial distribution pattern (Fig. 4A, C-Profiles 1 and 2). Interestingly, large urbanised areas like Naples and Salerno have a low abundance of volcanic rock elements despite being located close to large volcanic centres (Fig. 4A, C-Profiles 1 and 2). The reason for this unexpected finding is that these city soils have different contamination sources (e.g., heavy traffic, industrial centres) that might affect the spatial pattern. Profile 2 reveals a SW to NE decrease in volcanic rock elements extending from the volcanic centres towards the siliciclastic zone of the southern Apennines (Fig. 4C).

4.2.2.2. Z6 (ilr-6) balance map. The Z6 balance map represents the spatial abundance of Terra Rossa soil and Siliciclastic rock elements (Group B: Cr, Ni, Co and Cd) compared to other PTEs that primarily have anthropogenic contamination sources (Groups A and D) (Fig. 4B, Table 2). The range of balance values are quite symmetric (−2.33–2.71) indicating that no element group is dominant between Group B and Groups A and D. The negative balance values indicate that the spatial abundance of Terra Rossa soil and Siliciclastic rock elements (Group B) exceeds that of elements belonging to Groups A and D (Fig. 4B).

The highest spatial abundance of Group B (−2.33–1.7) can be observed in the eastern part of the Campania Region where siliciclastic rocks including shaley facies of the Crete Nere Fm. are dominant. The presence of Co, Cr and Ni was attributed to siliciclastic rocks and alteration from weathering and co-precipitation processes by Albanese et al. (2007) and Buccianti et al. (2015). The other high spatial dominance of Group B can be observed overlapping some of the large carbonate massifs where Terra Rossa soils (e.g., Haplic Luvisols) with bauxite mineralisation (Mondillo et al., 2011) are present (Fig. 4B). High concentrations of Co, Cr, Ni and Cd have also been observed in Terra Rossa soils located in Slovenia and Croatia (Durn et al., 1999,

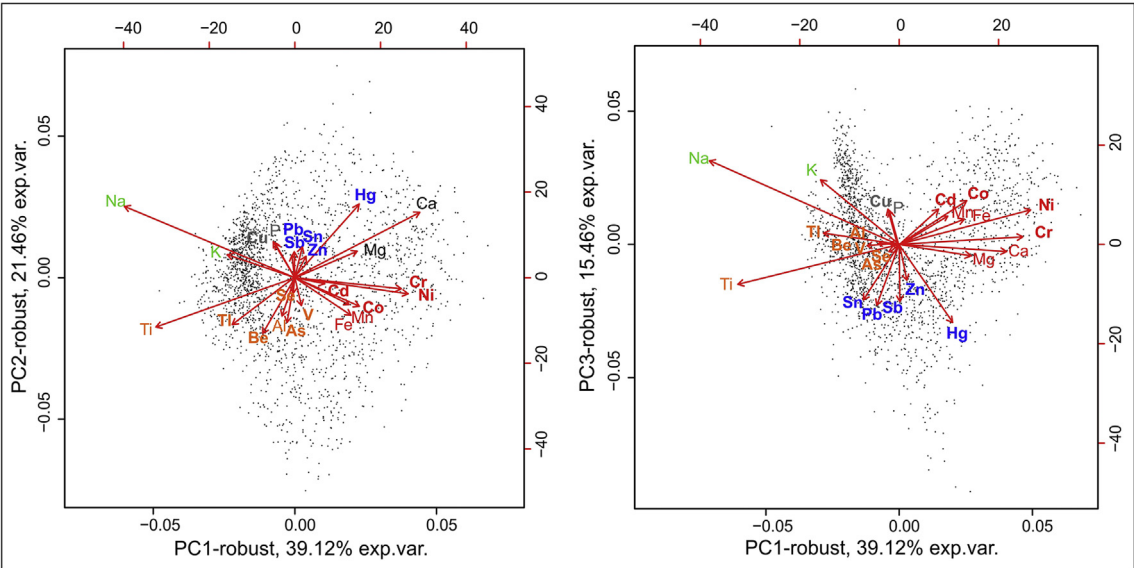


Fig. 3. Robust biplots of the 15 PTEs and 9 major elements projected using (A) PC1-PC2 axes and (B) PC1-PC3 axes. Elements with bold letters represent the 15 PTEs. Different colours mark different subgroups of elements. Further details about the separated subgroups can be found in the text. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2

Sequential binary partition table of the 15 PTEs subcomposition to obtain balances (Z1-Z14). Positive and negative signs (parts) in the i-th order partition show the separated two parts (groups of elements) used in the calculation of the given balance. The empty entries are parts (elements) that were not involved in the partition in the ith order. Balances were calculated using Eq. (1) with Co-DAPack 2.02.21. The highlighted balances were further analysed for the identification of spatial patterns.

PTEs	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14
As	+	+	+	+										
Be	+	+	+	-										
Tl	+	+	-											
V	+	-			+									
Se	+	-			-									
Pb	-					+	-	-	-	-	-			
Sb	-					+	-	-	-	-	+			
Sn	-					+	-	-	-	+				
Zn	-					+	-	-	+					
Hg	-					+	-	+						
Cu	-					+	+							
Cr	-											+	+	
Ni	-											+	-	
Co	-											-		+
Cd	-											-		-

2001; Miko et al., 1999).

There is a distinct separation between high positive balance values associated with elements that mainly have an anthropogenic origin in the Campania Plain area (Groups A and D) and the rest of the region where urbanisation is less dominant (Figs. 1B and 4B). Profiles 3 and 4 clearly show the decrease in balances extending away from the highly urbanised Campania Plain area towards the southern Apennines (Fig. 4C).

4.2.2.3. Z7 (ilr-7) balance map. The Z7 balance map reveals the spatial dominance of Cu compared to other PTEs that belong to Group A (Fig. 5A, Table 2). The positive balance values show the abundance of Cu with respect to the other elements. The highest spatial abundance of Cu is related to the slopes of Mt. Somma-Vesuvius and the Sarno River Basin where expansive agricultural areas utilize fertilisers and sewage sludge containing Cucompounds (Fig. 1B). High Cu concentrations are observed in scattered patches in the eastern part of the Campania

Region around Ariano Irpino where there are numerous olive orchards and vineyards (Fig. 1B). Cicchella et al. (2016) also emphasised that intensive agriculture in the Sarno River Basin is a primary contamination source for Cu and P.

Profile 5 clearly shows large peaks (higher spatial abundance of Cu) in the Sarno River Basin and reveals two large depressions in highly urbanised regions (Naples and Salerno) where other anthropogenic elements of Group A are abundant (Fig. 5C). Profile 6 unveils a large break in the data northeast of Salerno and a progressive decrease in Z7 balance values toward Avellino, suggesting that elements of other anthropogenic origins (Group A) have become more dominant (Fig. 5C).

4.2.2.4. Z8 (ilr-8) balance map. The Z8 balance map compares the spatial abundance of Hg with other elements of Group A (Fig. 5B, Table 2). The negative balance values shown in Fig. 5B indicate that other elements of Group A (Pb, Sb, Sn and Zn) have a higher abundance than Hg. The highest dominance of Hg (−3.26–4.12) is not spatially coherent with other data suggesting that there are multiple processes governing the concentration of Hg, which was also confirmed by the robust biplot analysis (Fig. 3, Tables 1 and 2). The high spatial abundance of Hg is associated with the most urbanised areas between Naples and Caserta (Figs. 1B and 5B), where there is intensive vehicle traffic and oil refineries are present. The high Hg dominance can be seen in the Phlegraean Fields and the island of Ischia, and is attributed to hydrothermal activity of volcanic centres (Tarzia et al., 2002; Lima et al., 2003b; De Vivo et al., 2016). High Hg abundances are also observed in the southwestern part of the Campania Region, where there are outcrops of black shales known to contain higher concentrations of Hg (~0.18 mg/kg; Reimann and de Caritat, 1998). Other high abundances of Hg might be sourced from dense fault networks (0.4–0.6 km/km²) and related hydrothermal activity close to the Mt. Matese, where the highest precipitation rates also occur (1500–2000 mm/y). High precipitation rates and the formation of organic-rich topsoils were considered to be critical factors for the accumulation of Hg in Norway (Ottesen et al., 2013).

High abundances of other elements of Group A (Pb, Sb, Sn and Zn) are represented by large negative balance values (−5.03–5.91) occurring between Salerno and Avellino, where the large Solofra industrial district is located (Fig. 5B). High negative balances can also be observed around the large volcanic complexes of Roccamonfina and Mt. Somma-

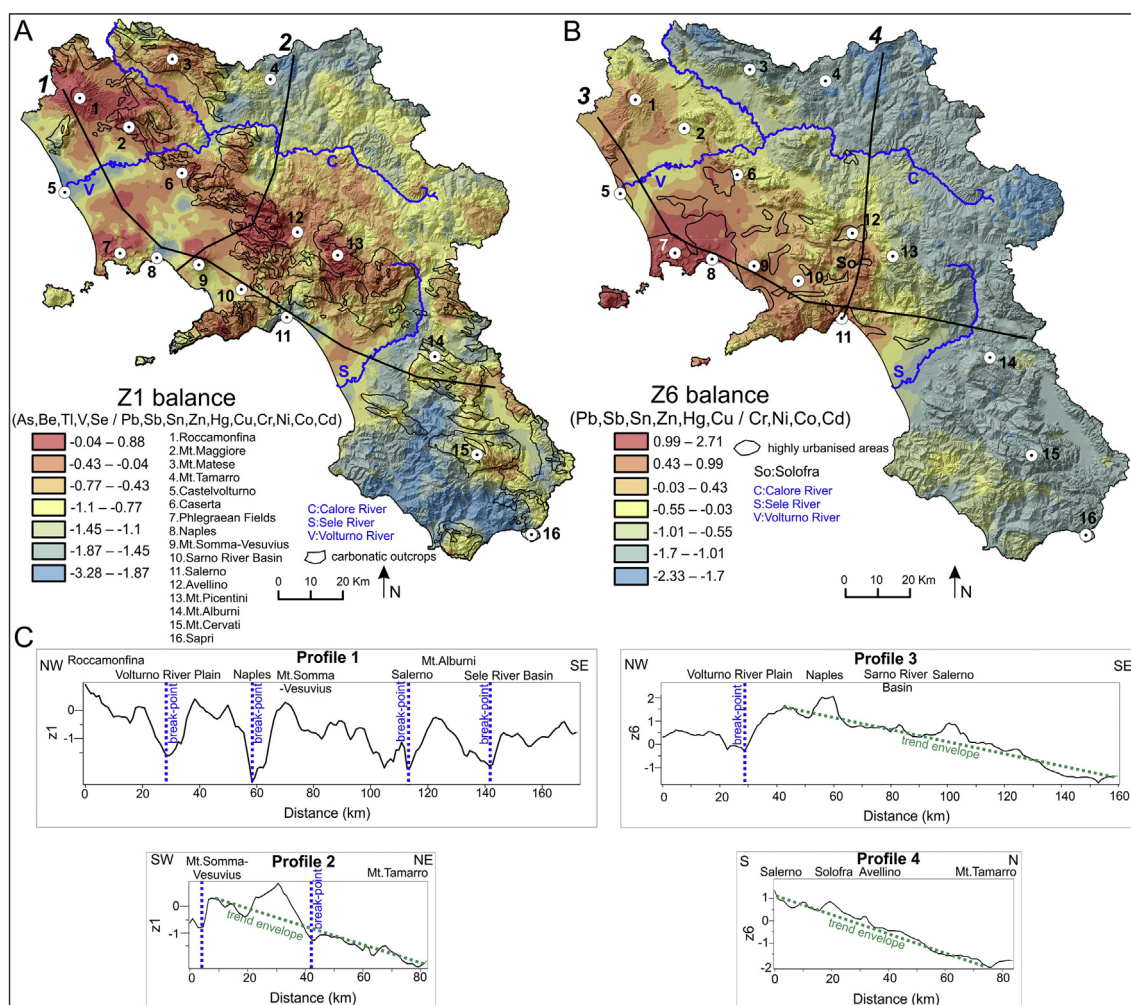


Fig. 4. (A) Z1 balance map showing the spatial abundance of volcanic rock elements (Group C: As, Be, Ti, V and Se) compared to other PTEs. (B) Z6 balance map demonstrating the spatial abundance of Terra Rossa soil and Siliciclastic rock elements (Group B: Cr, Ni, Co and Cd) compared to Group A (Pb, Sb, Sn, Zn and Hg) and Group D (Cu). (C) Profiles made in directions relevant for capturing the spatial variability and trend of abundances in different subgroups of the elements. The profile numbers and lines are shown on the maps (A and B). The sequential binary partition of the given balances can be seen in Table 2.

Vesuvius where, besides some local anthropogenic sources, geogenic processes (e.g., magmatic differentiation) control the enrichment of incompatible elements and PTEs in topsoil formed from volcanic rocks (Giannetti and Luhr, 1983; Ayuso et al., 1998; Peccerillo, 2005).

4.3. Compositional abundance maps and SCCI

In order to assess the level of contamination in the Campania Region, contamination abundance indices were calculated for the 15 PTEs and subgroups using Eq. (2).

The Compositional Abundance Index (CAI) can be interpreted as the relative difference in the spatial dominance of the PTEs, so it reflects the different amounts of enrichment or depletion in the study area. In addition, the calculation of the new SCCI (Soil Contamination Compositional Index) using Eq. (3) made it possible to quantify the contamination level of PTEs and reveal the contamination source patterns of different subgroups.

4.3.1. Total Compositional Abundance Index map

The Total Compositional Abundance Index map shows that the highest spatial abundance of PTEs (54.43–82.4) occurs in the highly urbanised areas of Naples and Salerno, where there is the highest concentration of heavy traffic and industrial activity (Fig. 6A). The high spatial abundance of PTEs is also associated with the Sarno River Basin,

which is one of the highest contaminated river basins in Italy due to a high level of industrial and agricultural activity (Cicchella et al., 2016) (Fig. 6A). Adamo et al. (2014) emphasised that even if the concentrations of some of the PTEs exceed the intervention limits, the bioavailability of these elements is low meaning that there are minimal risks for humans in the Sarno River Basin. The industrial district of Solofra between Salerno and Avellino also has a high spatial abundance of PTEs (54.43–82.4) related to the tannery industry, which represents one of the main contamination sources in the area (Fig. 6A).

The highest SCCI values (1.25–1.92) suggest that the above mentioned areas are highly contaminated (Fig. 7A). Indeed, the SCCI values indicate that the geometric means of the 15 PTEs can be almost twice as high as those of the respective baseline values in these areas (SCCI: 1.51–1.92, Fig. 7A). The heavy urbanisation, high industrial activity and dense road networks that define this part of the Campania Region are the sources for these results.

In contrast, other high spatial abundances of PTEs (54.43–82.4) in the Mt. Matese and the area between Sapri and Mt. Alburni are mainly from geogenic sources (e.g., Terra Rossa soils) and pyroclastic material (Fig. 6A). However, low SCCI values (1.001–1.12) reveal no contamination in these areas except Sapri where the values are medium-high (1.12–1.51) (Fig. 7A). Thiombane et al. (2017) suggested that the anthropogenic sources for Pb and Sb around urban areas like Sapri are frequent traffic jams.

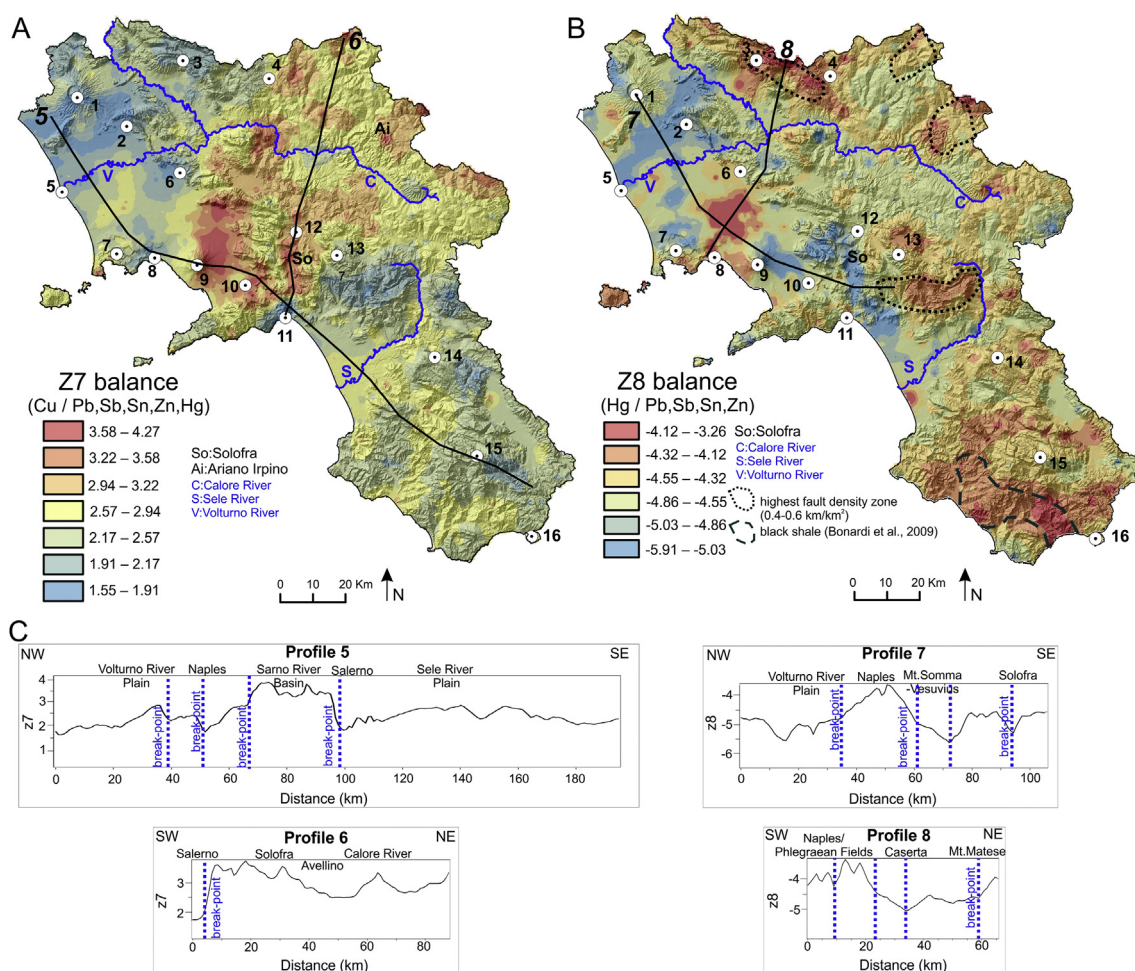


Fig. 5. (A) Z7 balance map showing the spatial abundance of Cu (Group D) compared to elements of Group A (Pb, Sb, Sn, Zn and Hg). (B) Z8 balance map demonstrating the spatial abundance of Hg with respect to other elements belonging to Group A (Pb, Sb, Sn and Zn). (C) Profiles made in directions relevant for capturing the spatial variability and trend of abundances in different subgroups of the elements. The profile numbers and lines are shown on the maps (A and B). The sequential binary partition of the given balances can be seen in Table 2. The topographic names indicated by 1–16 are the same as in Fig. 4A.

A low spatial dominance of PTEs (< 21.39) was identified in the eastern and southwestern parts of the Campania Region where urbanisation and industrial activity is rather low (Fig. 6A). Some medium-high SCCI values (1.12–1.51) were also revealed in the middle of the low abundance zone of PTEs around Ariano Irpino, where moderate contamination of the soils is possibly related to agricultural activity.

4.3.2. Volcanic Compositional Abundance Index map

The Volcanic Compositional Abundance Index map for Group C (As, Be, Se, Tl and V) shows the spatial abundance of volcanic rock elements (Fig. 6B). The highest dominance (60.8–86.42) is related to large volcanic centres, especially in the eastern and southern parts of Mt. Somma-Vesuvius (Fig. 6B). Petrik et al. (2018b and c) revealed that there were high concentrations of Be and As in topsoils over pyroclastic rocks related to the Ottaviano and Avellino eruptions (Ayuso et al., 1998). The spatial extent of these two eruptions corresponds to the high spatial abundance of volcanic elements. The Volcanic Compositional Abundance Index map also reveals that the carbonate massifs overlain by pyroclastic deposits have a high spatial abundance (> 60.8) of volcanic elements (Fig. 6B).

The SCCI values of Group C are very low and close to 1 in most of the study area indicating that there is little to no contamination by these elements (Fig. 7B). The lowest SCCI values (1.0007–1.006) found around large metropolitan areas (e.g., Naples, Salerno, Avellino) confirm that there are no anthropogenic contamination sources for these

elements. Slightly elevated SCCI values (1.07–1.13) from Mt. Somma-Vesuvius and the Phlegraean Fields can be explained by the presence of various volcanic soils types over different pyroclastic levels. The spatial coincidence of the highest abundance zones on the Volcanic and Total Compositional Abundance Index maps confirms that volcanic elements (Group A) certainly contribute to the high dominance of the 15 PTEs, especially around large volcanic centres (Fig. 6A and B). According to Albanese (2008), the bioavailability of these elements is less than 10% for urban soils. The lowest abundance of volcanic rock elements (< 27.91) is in the northeastern and southwestern parts of the Campania Region where siliciclastic rocks are found without pyroclastic deposits (Fig. 6B).

4.3.3. Terra Rossa soil and Siliciclastic Compositional Abundance Index map

This map reveals the spatial pattern and abundance of Terra Rossa soils and Siliciclastic rock elements (Group B: Co, Cr, Ni and Cd). The highest spatial abundance (58.11–84.17) is related to the carbonate massifs of the southern Apennines (e.g., Mt. Matese and Mt. Alburni) and in the area from Sapri to Mt. Cervati where Terra Rossa soils (e.g., Haplic Luvisols) with evidence for bauxite mineralisation can be found (Fig. 6C). The high dominance can also be observed east of the Calore River where siliciclastic rocks with shaley facies are present (e.g., Crete Nere Fm.) (Fig. 6C). However, these areas are characterised by very low SCCI values (< 1.01) that typically signify there are no sources of

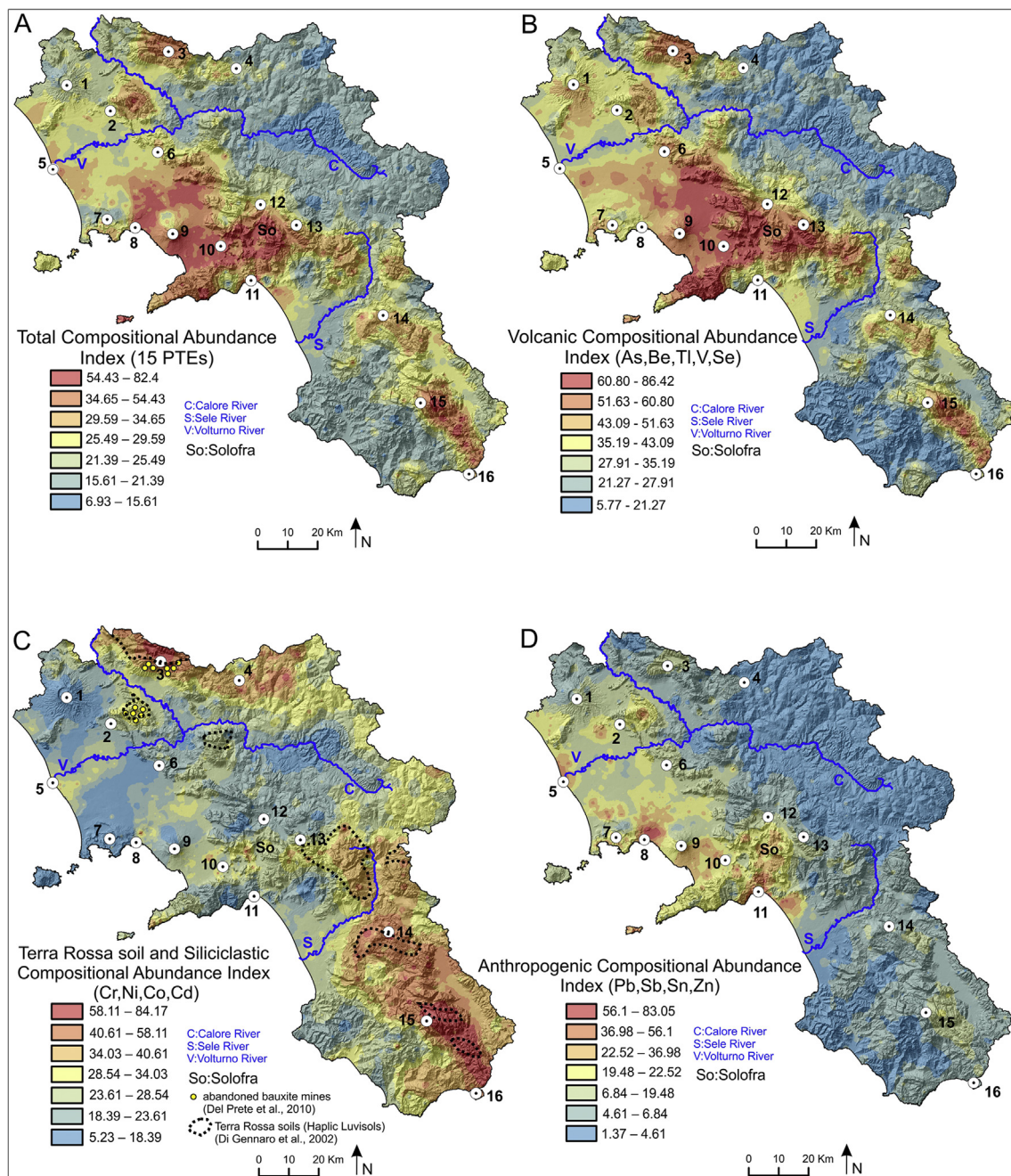


Fig. 6. (A) Total Compositional Abundance Index map showing the spatial abundance of the 15 PTEs. (B) Volcanic Compositional Abundance Index map illustrating the spatial dominance of volcanic elements (Group C). (C) Terra Rossa soil and Siliciclastic Compositional Abundance Index map demonstrating the spatial dominance of Group B elements (Cr, Ni, Co and Cd). (D) Anthropogenic Compositional Abundance Index map showing the spatial dominance of elements with anthropogenic sources (Group A).

contamination (Fig. 7C). Cicchella et al. (2005) and Albanese et al. (2007) suggested that lithology (e.g., siliciclastic rocks) can play an important role in controlling the accumulation of these elements.

However, high spatial abundance zones (> 58.11) were also found around large urbanised areas (e.g., Naples), industrial zones (e.g., Solofra) and at the mouths of the Sele and Volturno Rivers (Fig. 6C). The highest SCCI values (1.29–1.82) reveal that these areas are highly contaminated by elements of Group B (Fig. 7C). These areas are characterised by high industrial activity (e.g., tannery industry in Solofra Basin: Cr and Ni), intensive agricultural activity (e.g., Sarno River Basin: Cd, Co and Cr) and heavy vehicle traffic (Cd). Extensive cultivation occurs along alluvial plains that border large rivers, but in the case of the Volturno River Plain, the role of co-precipitation induced by

Fe- and Mn-oxyhydroxides in soils of alluvial origin cannot be ruled out (Ducci et al., 2017).

4.3.4. Anthropogenic Compositional Abundance Index map

Finally, we analysed the spatial abundance of elements that clearly have an anthropogenic origin (Group A: Pb, Sb, Sn and Zn). The Anthropogenic Compositional Abundance Index map unveils the highest spatial dominance (56.1–83.05) around highly urbanised areas like Naples and Salerno and in areas with extensive agricultural and industrial activity such as the Sarno River Basin and Solofra industrial district, respectively (Fig. 6D). The highest levels of contamination (SCCI: 1.58–1.88) occur around large metropolitan areas where the geometric means of the 15 PTEs can be twice as high as the counterpart

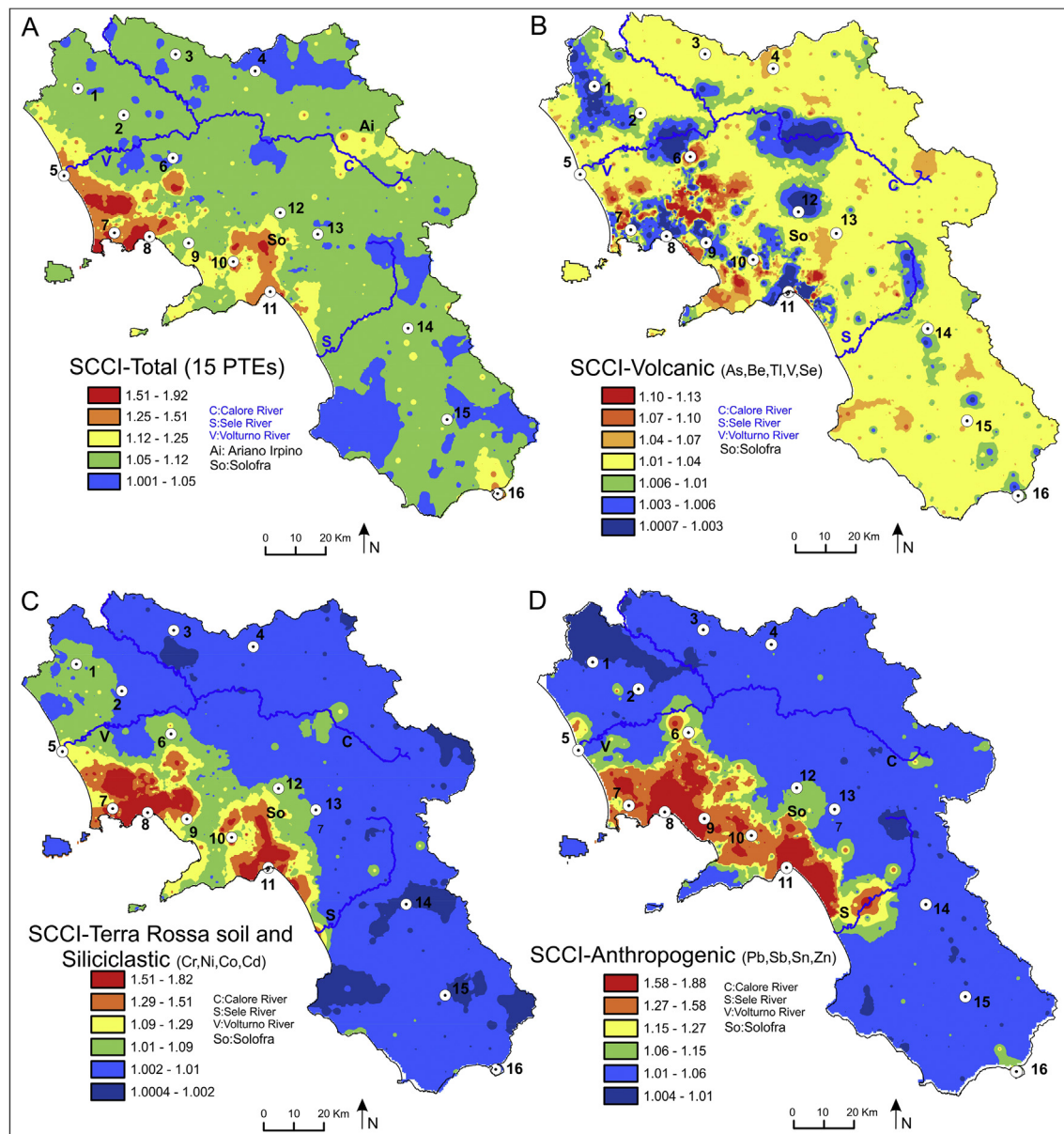


Fig. 7. (A) Soil Contamination Compositional Index (SCCI) map of the 15 PTEs. (B) The SCCI map of the Volcanic rock elements (As, Be, Se, Ti and V). (C) The SCCI map of the Terra Rossa soil and Siliciclastic elements (Cr, Ni, Co and Cd). (D) The SCCI map of the Anthropogenic elements (Pb, Sb, Sn and Zn). The SCCI was calculated using Eq. (3). The topographic names indicated by 1–16 are the same as in Fig. 4A.

baseline values (Fig. 7D). The primary role of the anthropogenic sources of Pb and Zn in this area was also described by Petrik et al. (2018a). Some sites around Sapri have moderately high SCCI values (1.15–1.27), which likely reflect contributions from heavy traffic in the urbanised centre (Fig. 7D). However, the geogenic contribution of pyroclastics from volcanic centres should also be considered as a source since elements of Group A (Pb, Sb, Sn and Zn) are also enriched from volcanic processes (Peccerillo, 2005).

The low dominance of the Anthropogenic Compositional Abundance Index (< 19.48) was found in the eastern and southwestern parts of the Campania Region away from highly urbanised and industrial areas (Fig. 6D). These areas have very low SCCI values associated with Group A (< 1.06) and therefore indicate that there is no anthropogenic contamination (Fig. 7D).

5. Conclusions

This study focuses on providing a better understanding of the

compositional behaviour and relative structure of the 15 PTEs in the Campania Region, which is considered to be a highly contaminated region in Italy. Special emphasis was put on the relative abundance and ratios of specific subcompositions (balances) of PTEs by mapping spatial patterns and making profiles to reveal spatial variability.

Preliminary steps are provided for analysing the spatial abundance of PTEs by applying a compositional abundance index. In addition, a new soil contamination compositional index was elaborated to quantify the contamination of topsoil by the 15 PTEs and subgroups.

The high spatial abundance of the 15 PTEs is related to highly urbanised (Naples and Salerno), highly industrial (Solofra) and intensive agricultural areas (Sarno River Basin), where the anthropogenic subgroup (Pb, Sb, Sn and Zn) has a high dominance and the high SCCI values indicate that contamination is from anthropogenic sources. The high spatial dominance of the volcanic rock subgroup (As, Be, Se, Ti and V) in the same area is certainly related to geogenic factors associated with the presence of pyroclastic deposits. Although the high spatial abundances of Group B elements (Cd, Cr, Co and Ni) are related to Terra

Rossa soils and shaley facies of siliciclastic rocks of the southern Apennines, these elements can also reach high abundances and reflect contamination (i.e. high SCCI values) in urbanised (e.g., traffic) and industrialised areas (e.g., tanneries, alloy manufacturing). Other high spatial abundances of the 15 PTEs with little or no contamination (i.e. very low SCCI values) are associated with carbonate massifs that reflect a mixture of geogenic factors (weathering, advanced pedogenic processes, adsorption and co-precipitation with Fe-/Mn-oxyhydroxides, presence of pyroclastic deposits) that might all be responsible for PTE accumulation.

The lowest spatial dominance of the 15 PTEs occurs in the north-eastern and southwestern siliciclastic zones of the Campania Region where urbanisation and industrial activity is minimal and therefore contamination by any source can be excluded.

A future goal is to improve the Compositional Abundance Index and the Soil Contamination Compositional Index by giving weights to the PTEs based on the bioavailability and toxicity levels of each element.

Acknowledgements

We thank the two anonymous reviewers for comments and suggestions that helped improve our paper.

This study was made possible by the financial support of the MIUR through two industry projects: 1) “Integrated agro-industrial chains with high energy efficiency for the development of eco-compatible processes of energy and biochemicals production from renewable sources and for the land valorization (EnerbioChem)” PON01_01966, funded within the framework of Operative National Program Research and Competitiveness 2007–2013 D. D. Prot. n. 01/Ric. 18.1.2010; and 2) “Development of green technologies for production of BIOchemicals and their use in preparation and industrial application of POLimeric materials from agricultural biomasses cultivated in a sustainable way in Campania Region (BioPoliS)” PON03PE_00107_1, funded within the framework of Operative National Program Research and Competitiveness 2007–2013 D. D. Prot. N. 713/Ric. 29/10/2010” (Research Units Responsible: Prof. B. De Vivo). This study has also been supported by the Istituto Zooprofilattico Sperimentale del Mezzogiorno (IZSM) by means of the Contract with the Centro Interdipartimentale di Ricerca Ambientale (CIRAM) within the framework of the “Campania Trasparente - Attività di monitoraggio integrato per la Regione Campania” project funded within the framework of Del.G.R. n. 497/2013: Fondo per le Misure Anticicliche e la Salvaguardia dell’Occupazione - Azione B4 “Mappatura del Territorio” approved with Executive Decree DG “Sviluppo Economico” n.585, 14/09/2015 (Research Unit Responsible: Prof. B. De Vivo).

References

- Adamo, P., Zampella, M., 2008. Chemical speciation to assess potentially toxic metals (PTMs) bioavailability and geochemical forms in polluted soils. In: De Vivo, B., Belkin, H.E., Lima, A. (Eds.), *Environmental Geochemistry: Site Characterization, Data Analysis and Case Histories*. Elsevier, pp. 175–212.
- Adamo, P., Iavazzo, P., Albanese, S., Agrelli, D., De Vivo, B., Lima, A., 2014. Bioavailability and soil-to-plant transfer factors as indicators of potentially toxic element contamination in agricultural soils. *Sci. Total Environ.* 500–501, 11–22.
- Aitchison, J., 1986. *The Statistical Analysis of Compositional Data*. Chapman & Hall, London, pp. 416.
- Aitchison, J., Greenacre, M., 2002. Biplots of compositional data. *J. Roy. Statist. Soc., C (Appl. Statist.)* 51 (4), 375–392.
- Aitchison, J., Egozcue, J., 2005. Compositional data analysis: where are we and where should we be heading? *Math. Geol.* 37, 829–850. <https://doi.org/10.1007/s11004-005-7383-7>.
- Albanese, S., 2008. Evaluation of the bioavailability of potentially harmful elements in urban soils through ammonium acetate-EDTA extraction: a case study in southern Italy. *Geochem. Explor. Environ. Anal.* 8, 49–57.
- Albanese, S., De Vivo, B., Lima, A., Cicchella, D., 2007. Geochemical background and baseline values of toxic elements in stream sediments of Campania region (Italy). *J. Geochem. Explor.* 93 (1), 21–34.
- Albanese, S., De Vivo, B., Lima, A., Cicchella, D., Civitillo, D., Cosenza, A., 2010. Geochemical baselines and risk assessment of the Bagnoli brownfield site coastal sea sediments (Naples, Italy). *J. Geochem. Explor.* 105, 19–33.
- Albanese, S., Sadeghi, M., Lima, A., Cicchella, D., Dinelli, E., Valera, P., Falconi, M., Demetriades, A., De Vivo, B., The GEMAS project team, 2015. GEMAS: Cobalt, Cr and Ni distribution in agricultural and grazing land soil of Europe. *J. Geochem. Explor.* 154, 81–93.
- Ayuso, R.A., De Vivo, B., Rolandi, G., Seal II, R.R., Paone, A., 1998. Geochemical and isotopic (Nd-Pb-Sr-O) variations bearing on the genesis of volcanic rocks from Vesuvius. *Italy. J. of Volc. Geoth. Res.* 82, 53–78.
- Barkouch, Y., Sedki, A., Ponesu, A., 2007. A new approach for understanding lead transfer in agricultural soil. *Water, Air, Soil Pollut.* 186 (1–4), 3–13. <https://doi.org/10.1007/s11270-007-9450-9>.
- Björklund, A., Gustavsson, N., 1987. Visualization of geochemical data on maps: new options. *J. Geochem. Explor.* 29, 89–103.
- Bonardi, G., Ciarcia, S., Di Nocera, S., Matano, F., Sgroso, I., Torre, M., 2009. Carta delle principali unità cinematiche dell’Appennino meridionale. Nota illustrativa. *Italian J. of Geosci.* 128 (1), 47–60.
- Boni, M., Rollinson, G., Mondillo, N., Balassone, G., Santoro, L., 2013. Quantitative mineralogical characterization of karst bauxite deposits in the southern Apennines. *Italy. Econ. Geol.* 108, 813–833.
- Buccianti, A., Lima, A., Albanese, S., Cannatelli, C., Esposito, R., De Vivo, B., 2015. Exploring topsoil geochemistry from the CoDA (compositional data analysis) perspective: the multi-element data archive of the Campania region (southern Italy). *J. Geochem. Explor.* 159, 302–316.
- Bureau Veritas Minerals, 2017. AQ250 (Ultra Trace Geochemical Aqua Regia Digestion), vols. 1–2.
- Chayes, F., 1960. On correlation between variables of constant sum. *J. Geophys. Res.* 65 (12), 4185–4193.
- Cheng, Q., 1999. Multifractality and spatial statistics. *Comput. Geosci.* 25, 949–961.
- Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. *J. Geochem. Explor.* 51, 109–130.
- Cheng, Q., Xu, Y., Grunsky, E., 2000. Integrated spatial and spectrum method for geochemical anomaly separation. *Nat. Resour. Res.* 9, 43–56.
- Chester, R., Stoner, J.H., 1973. Pb in particulates from the lower atmosphere of the eastern Atlantic. *Nature* 245, 27–28.
- Cicchella, D., De Vivo, B., Lima, A., 2005. Background and baseline concentration values of elements harmful to human health in the volcanic soils of the metropolitan and provincial area of Napoli (Italy). *Geochem. Explor. Environ. Anal.* 5, 29–40.
- Cicchella, D., Hoogewerff, J., Albanese, S., Adamo, P., Lima, A., Taiani, M.V.E., De Vivo, B., 2016. Distribution of toxic elements and transfer from the environment to humans traced by using lead isotopes. A case study in the Sarno River basin, south Italy. *Environ. Geochem. Health* 38, 619–637.
- Comas-Cufi, M., Thió-Henestrosa, S., 2011. CoDaPack 2.0: a stand-alone, multi-platform compositional software. In: Egozcue, J.J., Tolosana-Delgado, R., Ortego, M.I. (Eds.), *CoDaWork’11: 4th International Workshop on Compositional Data Analysis*. SantFeliu de Guixols.
- De Vivo, B., Rolandi, G., Gans, P.B., Calvert, A., Bohrsen, W.A., Spera, F.J., Belkin, H.E., 2001. New constraints on the pyroclastic eruptive history of the Campanian volcanic Plain (Italy). *Mineral. Petrol.* 73/ (1–3), 47–65 ISO 690.
- De Vivo, B., Petrosino, P., Lima, A., Rolandi, G., Belkin, H.E., 2010. Research progress in volcanology in Neapolitan area, Southern Italy: a review and alternative views. *Mineral. Petrol.* 99, 1–28.
- De Vivo, B., Lima, A., Albanese, S., Cicchella, D., Rezza, C., Civitillo, D., Minolfi, G., Zuzolo, D., 2016. Atlante geochimico-ambientale dei suoli della Campania (Environmental Geochemical Atlas of Campania Soils). Aracne Editrice, Roma, 978-88-548-9744-1pp. 364 (in Italian).
- Di Gennaro, A., Aronne, G., De Mascellis, R., Vingiani, S., Sarnataro, M., Abalsamo, P., Cona, F., Vitelli, L., Arpaia, G., 2002. I sistemi di terre della Campania. Monografia e carta 1:250 000, con legenda. Istituto per i sistemi agricoli e forestali del mediterraneo (ISAFoM).
- Ducci, D., Albanese, S., Boccia, L., Celentano, E., Cervelli, E., Corniello, A., Crispo, A., De Vivo, B., Iodice, P., Langella, C., Lima, A., Manno, M., Palladino, M., Pindozzi, S., Rigillo, M., Romano, N., Sellerino, M., Senatore, A., Speranza, G., Fiorentino, N., Fagnano, M., 2017. An integrated approach for the environmental characterization of a wide potentially contaminated area in southern Italy. *Int. J. Environ. Res. Publ. Health* 14, 1–23.
- Durn, G., Ottner, F., Slovenec, D., 1999. Mineralogical and geochemical indicators of the polygenetic nature of terra rossa in Istria, Croatia. *Geoderma* 91, 125–150.
- Durn, G., Slovenec, D., Čović, M., 2001. Distribution of iron and manganese in Terra Rossa from Istria and its genetic implications. *Geologica Croatica* 54, 27–36.
- Egozcue, J.J., 2009. Reply to “On the Harker variation diagrams. by J. A. Cortés. *Math. Geosci.* 41 (7), 829–834.
- Egozcue, J.J., Pawłowsky-Glahn, V., 2005. Groups of parts and their balances in compositional data analysis. *Math. Geol.* 37 (7), 795–828.
- Egozcue, J.J., Pawłowsky-Glahn, V., 2006. Simplicial geometry for compositional data. In: Buccianti, A.A., Mateu-Figueras, G., Pawłowsky-Glahn, V. (Eds.), *Compositional Data Analysis in the Geosciences. From Theory to Practice*, vol. 264. Geological Society, London, Special Publications, pp. 145–159.
- Egozcue, J.J., Pawłowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. *Math. Geol.* 35 (3), 279–300.
- Fabian, K., Reimann, C., Caritat, P. de, 2017. Quantifying diffuse contamination: method and application to Pb in soil. *Environ. Sci. Technol.* 51, 6719–6726.
- Filzmoser, P., Hron, K., 2008. Outlier detection for compositional data using robust methods. *Math. Geosci.* 40 (3), 233–248.
- Filzmoser, P., Hron, K., Reimann, C., 2009. Principal component analysis for compositional data with outliers. *Environmetrics* 20 (6), 621–632.
- Gabriel, K.R., 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58 (3), 453–467.
- Giannetti, B., Luhr, J.F., 1983. The white trachytic tuff of Roccamonfina volcano (roman region, Italy). *Contrib. Mineral. Petrol.* 84, 235–252.
- Guillén, M.T., Delgado, J., Albanese, S., Nieto, J.M., Lima, A., De Vivo, B., 2011. Environmental geochemical mapping of Huelva municipality soils (SW Spain) as a

- tool to determine background and baseline values. *J. Geochem. Explor.* 109 (1–3), 59–69. <https://doi.org/10.1016/j.gexplo.2011.03.003>.
- Hakanson, L., 1980. An ecological risk index for aquatic pollution control. A sedimentological approach. *Water Res.* 14 (8), 975–1001. [https://doi.org/10.1016/0043-1354\(80\)90143-8](https://doi.org/10.1016/0043-1354(80)90143-8).
- Han, J., Kamber, M., 2001. *Data Mining: Concepts and Techniques*. Morgan-Kaufmann Academic Press, San Francisco, pp. 740.
- Hawkes, H.E., Webb, J.S., 1962. *Geochemistry in Mineral Exploration*. Harper, New York, pp. 415.
- Ippolito, F., Ortolani, F., Russo, M., 1973. *Struttura marginale tirrenica dell'Appennino campano: reinterpretazione di dati di antiche ricerche di idrocarburi*. *Memor. Soc. Geol. Ital.* 12, 227–250.
- ISTAT, 2016. Resident Population on 1st January, 2017. <https://www.istat.it/en/population-and-597.households>.
- Jarauta-Bragulat, E., Hervada-Sala, C., Egozcue, J.J., 2015. Air quality index revisited from a compositional point of view. *Math. Geosci.* <https://doi.org/10.1007/s11004-015-9599-5>.
- Ji, Y., Feng, Y., Wu, J., Zhu, T., Bai, Z., Duan, C., 2008. Using geoaccumulation index to study source profiles of soil dust in China. *J. Environ. Sci.* 20 (5), 571–578. [https://doi.org/10.1016/S1001-0742\(08\)62096-3](https://doi.org/10.1016/S1001-0742(08)62096-3).
- Kabata-Pendias, A., 2011. *Trace Elements of Soils and Plants*, fourth ed. CRC Press, Taylor & Francis Group, LLC, USA, pp. 28–534.
- Kürzl, H., 1988. Exploratory data analysis: recent advances for the interpretation of geochemical data. *J. Geochem. Explor.* 30, 309–322.
- Lepeltier, C., 1969. A simplified statistical treatment of geochemical data by graphical representation. *Econ. Geol.* 64, 538–550.
- Lim, H.S., Lee, J.S., Chon, H.T., Sager, M., 2008. Heavy metal contamination and health risk assessment in the vicinity of the abandoned Songcheon Au–Ag mine in Korea. *J. Geochem. Explor.* 96 (2–3), 223–230. <https://doi.org/10.1016/j.gexplo.2007.04.008>.
- Lima, A., De Vivo, B., Cicchella, D., Cortini, M., Albanese, S., 2003a. Multifractal IDW interpolation and fractal filtering method in environmental studies: an application on regional stream sediments of Campania Region (Italy). *Appl. Geochem.* 18 (12), 1853–1865. [https://doi.org/10.1016/S0883-2927\(03\)00083-0](https://doi.org/10.1016/S0883-2927(03)00083-0).
- Lima, A., Cicchella, D., Di Francia, S., 2003b. Natural contribution of harmful elements in thermal groundwaters of Ischia island (southern Italy). *Environ. Geol.* 43, 930–940.
- Lima, A., Albanese, S., De Vivo, B., 2005. Geochemical baselines for the radioelements K, U and Th in the Campania region, Italy: a comparison of stream-sediment geochemistry and gamma-ray surveys. *Appl. Geochem.* 20, 611–625.
- McKinley, J.M., Hron, K., Grunsky, E.C., Reimann, C., Caritat, P. de, Filzmoser, P., van den Boogaart, K.G., Tolosana-Delgado, R., 2016. The single component geochemical map: fact or fiction? *J. Geochem. Explor.* 162, 16–28.
- Miko, S., Durn, G., Prohić, E., 1999. Evaluation of terra rossa geochemical baselines from Croatian karst regions. *J. Geochem. Explor.* 66, 173–182.
- Minolfi, G., Petrik, A., Albanese, S., Lima, A., Rezza, C., De Vivo, B., 2018. Lead, Cu and Zn distributions in topsoils of the Campania region, Italy. *Spec. Issue Geochem.: Explor. Environ. Anal.* <https://doi.org/10.1144/geochem2017-074>.
- Mondillo, N., Balassone, G., Boni, M., Rollinson, G., 2011. Karst bauxites in the Campania Apennines (southern Italy): a new approach. *Period. Mineral.* 80 (3), 407–432.
- Muller, G., 1969. Index of geoaccumulation in sediments of the rhine river. *GeoJournal* 2, 108–118.
- Muller, G., 1981. The heavy metal pollution of the sediments of Neckars and its tributary: a stocktaking. *Chem. Ztg.* 105, 157–164.
- Otero, N., Tolosana-Delgado, R., Solera, A., Pawlowsky-Glahn, V., Canals, A., 2005. Relative vs. absolute statistical analysis of compositions: a comparative study of surface waters of a Mediterranean river. *Water Res.* 39, 1404–1414.
- Ottesen, R.T., Birke, M., Finne, T.E., Gosar, M., Locutura, J., Reimann, C., Tarvainen, T., the GEMAS Project Team, 2013. Mercury in European agricultural and grazing land soils. *Appl. Geochem.* 33, 1–12.
- Palarea-Albaladejo, J., Martín-Fernández, J.A., 2013. Values below detection limit in compositional chemical data. *Anal. Chim. Acta* 764, 32–43.
- Parsa, M., Maghsoudi, A., Yousefi, M., Carranza, J.M., 2017. Multifractal interpolation and spectrum-area fractal modeling of stream sediment geochemical data: implications for mapping exploration targets. *J. Afr. Earth Sci.* 128, 5–15.
- Pawlowsky-Glahn, V., Buccianti, A., 2011. *Compositional Data Analysis: Theory and Applications*. John Wiley & Sons, pp. 400.
- Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R., 2015. *Modelling and Analysis of Compositional Data*. John Wiley & Sons, pp. 252.
- Peccerillo, A., 2005. Plio-quaternary volcanism in Italy. In: *Petrology, Geochemistry, Geodynamics*. Springer-Verlag, Berlin Heidelberg, pp. 365 ISBN 978-3-540-29,092-6.
- Petrik, A., Thiombane, M., Albanese, S., Lima, A., De Vivo, B., 2018a. Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) through a compositional discrimination analysis: a case study on the Campanian topsoil data. *Geoderma* 331, 87–99.
- Petrik, A., Albanese, S., Lima, A., De Vivo, B., 2018b. The spatial pattern of beryllium and its possible origin using compositional data analysis on a high-density topsoil data set from the Campania Region (Italy). *Appl. Geochem.* 91, 162–173. <https://doi.org/10.1016/j.apgeochem.2018.02.008>.
- Petrik, A., Albanese, S., Lima, A., Jordan, Gy, De Vivo, B., Rolandi, R., Rezza, C., 2018c. Spatial pattern recognition of arsenic in topsoil using high-density regional data. *Spec. Issue Geochem.: Explor. Environ. Anal.* <https://doi.org/10.1144/geochem2017-060>.
- Qu, C., Albanese, S., Lima, A., Li, J., Doherty, A.L., Qi, S., De Vivo, B., 2017. Residues of hexachlorobenzene and chlorinated cyclodien pesticides in the soils of the Campanian Plain, southern Italy. *Environ. Pol.* 231 (2), 1497–1506. <https://doi.org/10.1016/j.envpol.2017.08.100>.
- Reimann, C., Caritat, P. de, 1998. Chemical Elements in the Environment. *Factsheets for the Geochemist and Environmental Scientist*. Springer-Verlag Berlin, Heidelberg.
- Reimann, C., de Caritat, P., 2000. Intrinsic flaws of element enrichment factors (EFs) in environmental geochemistry. *Environ. Sci. Technol.* 34, 5084–5091.
- Reimann, C., de Caritat, P., 2005. Distinguishing between natural and anthropogenic sources for elements in the environment: regional geochemical surveys versus enrichment factors. *Sci. Total Environ.* 337, 91–107.
- Reimann, C., Garrett, R.G., 2005. Geochemical background – concept and reality. *Sci. Total Environ.* 350, 12–27.
- Reimann, C., Garrett, R.G., Filzmoser, P., 2005. Background and threshold – critical comparison of methods of determination. *Sci. Total Environ.* 346, 1–16.
- GEMAS Team, 2014. In: Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., O'Connor, P. (Eds.), *Chemistry of Europe's Agricultural Soils — Part a: Methodology and Interpretation of the GEMAS Data Set*. *Geologisches Jahrbuch (Reihe B)*, Schweizerbart: Hannover, pp. 528.
- Rezza, C., Petrik, A., Albanese, S., Lima, A., Minolfi, G., De Vivo, B., 2018. Mo, Sn and W patterns in topsoils of the Campania Region, Italy. *Spec. Issue Geochem.: Explor. Environ. Anal.* <https://doi.org/10.1144/geochem2017-061>.
- Rolandi, G., Petrosio, P., McGeehin, J., 1998. The interplinian activity at Somma Vesuvius in the last 3500 years. *J. Volcanol. Geoth. Res.* 82, 19–52.
- Rolandi, G., Bellucci, F., Heizler, M.T., Belkin, H.E., De Vivo, B., 2003. Tectonic controls on genesis of ignimbrites from the campanian volcanic zone, southern Italy. In: *In: De Vivo, B., Scandone, R. (Eds.), Ignimbrites of the Campania Plain, Italy*. *Mineral.Petrol.*, vol. 79. pp. 3–31.
- Rousseeuw, P.J., Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41, 212–223.
- Sinclair, A.J., 1974. Selection of threshold values in geochemical data using probability graphs. *J. Geochem. Explor.* 3, 129–149.
- Skordas, K., Kelepertsis, A., 2005. Soil contamination by toxic metals in the cultivated region of Agia, Thessaly, Greece. Identification of sources of contamination. *Environ. Geol.* 48 (4), 615–624. <https://doi.org/10.1007/s00254-005-1319-x>.
- Sucharova, J., Suchara, I., Hola, M., Marikova, S., Reimann, C., Boyd, R., Filzmoser, P., Englmaier, P., 2012. Top-/Bottom-soil ratios and enrichment factors: what do they really show? *Appl. Geochem.* 27, 138–145.
- Tarzia, M., De Vivo, B., Somma, R., Ayuso, R.A., McGill, R.A.R., Parrish, R.R., 2002. Anthropogenic versus natural pollution: an environmental study of an industrial site under remediation (Naples, Italy). *Geochem. Explor. Environ. Anal.* 2, 45–56.
- Tennant, C.B., White, M.L., 1959. Study of the distribution of some geochemical data. *Econ. Geol.* 54, 1281–1290.
- Thiombane, M., Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Cavaliere, M., De Vivo, B., 2017. Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model. *J. Geochem. Explor.* <https://doi.org/10.1016/j.gexplo.2017.06.010>.
- Tolosana-Delgado, R., van den Boogaart, K.G., 2013. Joint consistent mapping of high-dimensional geochemical surveys. *Math. Geosci.* 45, 983–1004.
- Torrente, M., Milia, A., 2013. Volcanism and faulting of the Campania margin (Eastern Tyrrhenian Sea, Italy): a three-dimensional visualization of a new volcanic field off Campi Flegrei. *Bull. Volcanol.* 75, 719. <https://doi.org/10.1007/s00445-013-0719-0>.
- Tukey, J.W., 1977. *Exploratory Data Analysis*. Addison-Wesley Publishing Company.
- Varrica, D., Tamburo, E., Milia, N., Vallascas, E., Cortimiglia, V., De Giudici, G., Dongarrà, G., Sanna, E., Monna, F., Losno, R., 2014. Metals and metalloids in hair samples of children living near the abandoned mine sites of Sulcis-Iglesiente (Sardinia, Italy). *Environ. Res.* 134 (Suppl. C), 366–374. <https://doi.org/10.1016/j.envres.2014.08.013>.
- Vercoutere, K., Fortunati, U., Muntau, H., Griepink, B., Maier, E.A., 1995. The certified reference materials CRM 142 R light sandy soil, CRM 143 R sewage sludge amended soil and CRM145R sewage sludge for quality control in monitoring environmental and soil pollution. *Fresen. J. Anal. Chem.* 352, 197–202.
- Vitale, S., Giacaria, S., 2013. Tectono-stratigraphic and kinematic evolution of the southern Apennines/Calabria-Peloritani Terrane system (Italy). *Tectonophysics* 583, 164–182.
- Wang, Y., Sikora, S., Kim, H., Dubey, B., Townsend, T., 2012. Mobilization of iron and arsenic from soil by construction and demolition debris landfill leachate. *Waste Manag.* 32 (5), 925–932. <https://doi.org/10.1016/j.wasman.2011.11.016>.
- Washington, H.S., 1906. *The Roman Comagmatic Region*, vol. 57. Carnegie Institution of Washington, pp. 199.
- Wedepohl, K.H., 1995. The composition of the continental crust. *Geochem. Cosmochim. Acta* 59, 1217–1232.
- Wu, S., Peng, S., Zhang, X., Wu, D., Luo, W., Zhang, T., Zhou, S., Yang, G., Wan, H., Wu, L., 2015. Levels and health risk assessments of heavy metals in urban soils in Dongguan, China. *J. Geochem. Explor.* 148, 71–78. <https://doi.org/10.1016/j.gexplo.2014.08.009>.
- Yang, J.Y., Yang, X.E., He, Z.L., Chen, G.C., Shentu, J.L., Li, T.Q., 2004. Adsorption-desorption characteristics of lead in variable charge soils. *J. Environ. Sci. and Health Part A* 39 (8), 1949–1967.
- Zhao, S., Feng, C., Wang, D., Liu, Y., Shen, Z., 2013. Salinity increases the mobility of Cd, Cu, Mn, and Pb in the sediments of Yangtze Estuary: relative role of sediments' properties and metal speciation. *Chemosphere* 91 (7), 977–984. <https://doi.org/10.1016/j.chemosphere.2013.02.001>.
- Zhao, L., Xu, Y., Hou, H., Shanguan, Y., Li, F., 2014. Source identification and health risk assessment of metals in urban soils around the Tanggu chemical industrial district, Tianjin, China. *Sci. Total Environ.* 468 (Suppl. C), 654–662. <https://doi.org/10.1016/j.scitotenv.2013.08.094>.
- Zuo, R., Wang, J., Chen, G., Yang, M., 2015. Identification of weak anomalies: a multifractal perspective. *J. Geochem. Explor.* 148, 12–24.
- Zuo, R., Wang, J., 2016. Fractal/multifractal modelling of geochemical data: a review. *J. Geochem. Explor.* 164, 33–41.

Paper 3

Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) through a compositional discrimination analysis: a case study on the Campanian topsoil data

Matar Thiombane, Attila Petrik, Stefano Albanese, Annamaria Lima, Benedetto De Vivo

Journal of Geoderma, Volume 331, 1st December 2018, Pages 87–99



Source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) through a compositional discrimination analysis: A case study on the Campanian topsoil data

Attila Petrik^{a,*,1}, Matar Thiombane^{a,1}, Stefano Albanese^a, Annamaria Lima^a, Benedetto De Vivo^{b,c}

^a Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse, Università degli Studi di Napoli "Federico II", Complesso Universitario Monte S. Angelo, Via Cintia snc, 80126 Naples, Italy

^b Pegaso University, Piazza Trieste e Trento 48, 80132 Napoli, Italy

^c BENECON Scrl, Dipartimento Ambiente e Territorio, Via S. Maria di Costantinopoli 104, 80138 Napoli, Italy

ARTICLE INFO

Handling Editor: Edward A Nater

Keywords:

Campania Region (Italy)
Soil geochemistry
Potentially toxic elements
Compositional data analysis
Discrimination index

ABSTRACT

One of the main objectives of environmental geochemistry is to reveal the source and spatial patterns of different inorganic and organic elements/compounds with special emphasis on potentially toxic elements. In the last couple of decades, environmental geochemists mainly focused on the identification and separation of geochemical anomalies from background/baseline values of potentially toxic heavy metals and the estimation of their ecological and human health risks. However, the main concern with previously published papers on this issue is that the majority of them simply neglected the compositional nature of geochemical data; hence results became spurious and biased and can be interpreted with reservations. Our goal is to identify, interpret and discriminate the source patterns of 4 potentially toxic elements in the Campania Region (Italy) emphasising on their ratios and spatial abundance using multivariate compositional data analysis. This study contributes to understand the compositional behaviour and relative proportion of 4 potentially toxic elements whose precise contamination sources have been unclear before. A workflow of compositional data analysis including a new discrimination index has been elaborated to identify their possible sources of contamination and enrichment. The investigated data set includes 18 elements derived from 3669 topsoil samples, collected at an average sampling density of 1 site per 3.2 km². First, robust biplots and factor analysis were performed to get an overview of elemental associations and reduce the dimensionality of the data set. They revealed that the 4 PTEs belong to different groups. The multivariate regression analysis using $\ln r$ -transformed (additive logratio) data proved the strong linear relationship between Fe, Mn (independent) and each investigated PTE (dependent), but also unveiled deviation trends in case of Zn and Pb. Based on the multivariate regression result, a sequential binary partition was performed by means of 6 variables (the 4 PTEs, Fe, and Mn) to obtain balances. Balances are $\ln r$ -coordinates (isometric-logratio) which can be interpreted as ratios of specific groups of elements. They were used to generate interpolated maps by using multifractal method to see spatial patterns and proportions of elemental associations. A new index has been elaborated based on the bivariate regression of balances and their standardised residuals, which was particularly useful to identify and separate the sources of anthropogenic contamination and geogenic enrichment of respective elemental associations. The large urban and industrial areas (e.g. Naples, Salerno) along the coastline are mainly contaminated by Pb and Zn due to heavy traffic and alloy production. Some Cr and Ni contamination was discerned in the Sarno Basin where the Solofra industrial district is likely to be the principal source through releases from tannery industry. The large volcanic complexes (e.g. Mt. Somma-Vesuvius, Phlegraean Fields, Mt. Roccamonfina) are all characterised by geogenic enrichment of Zn and Pb. In contrast, Cr and Ni-geogenic enrichment is mainly related to the siliciclastic deposits.

1. Introduction

Geochemical surveys aim to enhance mineralisation and

contamination through exploration and environmental geochemistry, respectively. For that, various geostatistical computations have been used to identify source patterns of different elements related to

* Corresponding author.

E-mail address: tectonics@caesar.elte.hu (A. Petrik).

¹ First authors with equal contribution.

underlying geological features and/or anthropogenic activities (Cheng et al., 1994; Cheng, 1999; Lima et al., 2003; Reimann and De Caritat, 2005). Graphical and statistical methods (e.g. classification using box-plots or cumulative probability plots) have been elaborated to study the elemental distribution and identify univariate outliers (Tennant and White, 1959; Sinclair, 1976, 1983; Kürzl, 1988; Reimann et al., 2008). The detection of multivariate outliers using certain cut-off values on the Mahalanobis distance and observed covariance ratio were also successfully applied on compositional data (Bollen, 1987; Roosseuw and Van Zomeren, 1990; Filzmoser et al., 2012; Buccianti et al., 2015).

Background/anomaly separation by using various fractal methods (e.g. Concentration-Area) was given particular attention in mineral exploration, environmental health-risk analysis and regional topsoil studies (Cheng et al., 1994; Cheng, 1999; Agterberg, 2001; Lima et al., 2003; Fabian et al., 2014; Albanese et al., 2015; Zuo et al., 2015). In addition, the question of anthropogenic or geogenic origin has always been in the interest of investigations. Several indices have been invented to separate anthropogenic contamination from geogenic background values like Enrichment Factor (EF, Chester and Stoner, 1973) or Geoaccumulation Index (Müller, 1979). These indices were critically reviewed by Reimann and de Caritat (2005) claiming that their values vary and depend on the different parent rock materials and chosen reference media as well as reference elements. In addition, these indices do not take into account the different biogeochemical processes which may have remarkable impact on elemental enrichment/contamination (Reimann and de Caritat, 2005). Finally, geochemical data are inherently compositional data and indices based on the absolute values of elements disregard the compositional principles like scaling invariance and subcompositional coherence (Pawłowsky-Glahn and Buccianti, 2011).

In this study, the main objective is to identify, interpret and map the source patterns of Zn, Pb, Cr and Ni potentially toxic elements (PTEs) on a high-density Campanian topsoil data using multivariate compositional data analysis. A new compositional discrimination index has also been elaborated taking into account the compositional nature of geochemical data to better understand and distinguish the precise contamination/enrichment sources of the 4 PTEs which have been unclear before.

2. Materials and methods

2.1. Survey area, sampling procedure and analytics

The Campania Region is located in the southern part of Italy, covering an area of about 13,600 km² (Fig. 1A). The region is bordered by the Tyrrhenian Sea from the west, the Basilicata Region from the south, the Apulia Region from the east, and Latium Region from the north (Fig. 1A).

The main geological features of the Campania Region are constituted by the NW-SE trending Apennine chain. The main parts of the Apennine chain are the Mt. Matese in the north, Mt. Camposauro, Mt. Avella and Mt. Picentini in the centre, and the Mt. Alburni and Mt. Cervati in the south (Fig. 1A). These mountains are mostly made up of sedimentary rocks such as limestone and dolostone whilst the external domains are constituted by siliceous schist and terrigenous sediments (clays, siltstone, sandstone, and conglomerate) (Bonardi et al., 1998; De Vivo et al., 2016) (Fig. 1A). Igneous rocks are also present in the region and are mostly formed by potassic and ultrapotassic volcanic rocks of the Mt. Roccamonfina in the northwest (Peccerillo, 2005), and the Mt. Somma-Vesuvius, Phlegraean Fields and Ischia volcanoes all along the western coastline (Rolandi et al., 2003; De Vivo et al., 2010, 2016) (Fig. 1A). The coastal areas and plains are built up of alluvial, lacustrine and coastal (mixes of oceanic and terrestrial) sediments (Bonardi et al., 1998; De Vivo et al., 2016).

The hydrography of the Campania Region is characterised by three main river catchments (Garigliano, Volturno and Sele) and by numerous minor streams, mostly flowing towards the Tyrrhenian Sea. The morphology of the drainage network is irregular and controlled by geological and structural features. Rivers are water source for irrigation of agricultural fields (Ducci and Tranfaglia, 2005).

Campania is one of the most populated regions of Italy with > 5.8 million inhabitants (ISTAT, 2016). This high density of population is coupled with the presence of a large number of industrial activities, with the majority being involved in agriculture: vineyards and olive plantations - mostly in hilly areas - seasonal crops, and greenhouse products represent major resources for the region and the local economy (Albanese et al., 2007). This intensive agriculture activity occupies > 50% of the total land and occurs mostly in the coastal and mountainous areas, where fertile land is available (Fig. 1B). Unfortunately, such agricultural activities and related industries (e.g. food processing) are known to have a potential negative impact - if not properly managed - on the contamination of natural resources such as

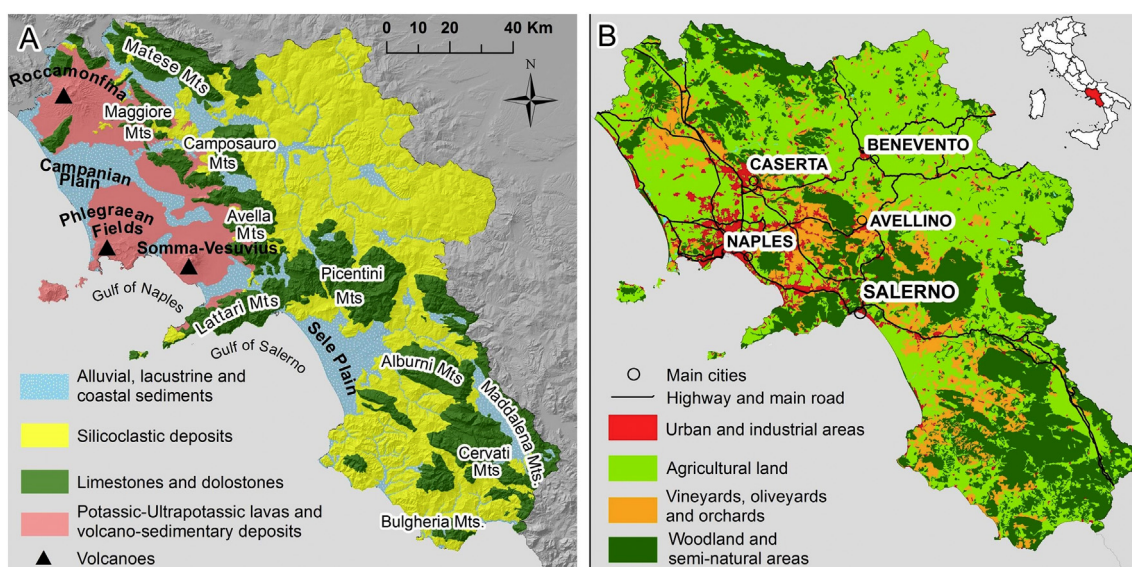


Fig. 1. Simplified geological map (A) and land use (B) of the Campania Region (Italy).

superficial and groundwater as well as soils. Campania is not immune to these problems, and some studies have already highlighted the harmful effects of anthropogenic contamination on natural resources (Cicchella et al., 2005; De Vivo et al., 2016; Minolfi et al., 2016).

From 2013 to 2015, 3669 samples were collected from topsoil of the Campania Region (13,600 km²) at a nominal density of 1 sample/3.2 km². Each top soil sample (from 0 to 20 cm) was obtained by homogenizing 5 subsamples at the corners and the centre of a 100m² square, collecting approximately 1.5 kg in total. The sampling procedure followed the Geochemical Mapping of Agricultural and Grazing Land Soils (GEMAS) sampling procedure described by Reimann et al. (2014). At each sampling site, several physico-chemical parameters of the soil properties were measured, including pH, total water content, conductivity, total organic content and the geographical coordinates recorded by geospatial positioning systems (GPS).

Chemical analyses were carried out at an international accredited laboratory, Acme Analytical Laboratories Ltd. (now Bureau Veritas, Vancouver, Canada). The samples were analysed after the ultratrace aqua regia extraction method (Bureau Veritas, 2017), by a combination of inductively coupled plasma atomic emission (ICP-AES) and inductively coupled plasma mass spectrometry (ICP/MS) for “pseudo-total” concentration of 53 elements. The “pseudototal” corresponds to the part concentration of elements extractable using strong acid solution (aqua regia digestion). It estimates the maximum amounts of metals that could hypothetically be mobilized and transported in the environment from geogenic part and all anthropogenic inputs (Vercoutere et al., 1995; Adamo and Zampella, 2008). Although primary silicates are not dissolved, metals associated with most other major soil components are liberated.

The 2-mm sieved soil was characterised according to the Italian official methods of soil analysis (Violante and Adamo, 2000). A subsample of 15 g was digested in 90 ml aqua regia and leached for 1 h in a 95 °C water bath. After cooling, the solution was diluted to a final volume of 300 ml using a solution of 5% HCl. The sample weight to solution volume ratio was 1 g per 20 ml. The solutions were analysed for the “pseudototal” concentration of 53 elements using a Perkin Elmer Elan 6000/9000 inductively coupled plasma emission mass spectrometer (ICP-MS). The accuracy and precision of the data was measured by comparison to known analytical standards. Calibration solutions were included at the beginning and the end of each analytical run (a total of 40 solutions). Precision is $\pm 100\%$ at the detection limit, and improves to better than $\pm 10\%$ at concentrations 50 times the detection limit or greater.

2.2. Compositional robust biplot and factor analysis

A workflow of compositional data analysis has been elaborated to obtain additional information to discriminate the sources of the 4 PTEs in the study area (Fig. 2).

To better visualize the element distributions and quantify possible natural or anthropogenic impact, robust compositional biplots were generated (Fig. 3). This is a powerful statistical tool that displays both samples and variables of a data matrix in terms of the resulting scores and loading (Gabriel, 1971). The scores represent the structure of the compositional data in the Euclidian space based on variance and covariance matrix; moreover they display the association structure of the dataset. Variables are represented by rays (or vectors) drawn from the centre of the plot, whose length is proportional to the amount of their explained variance (communality). The interpretation of the plot depends on the loading (the length of rays) structures and in more details on the approximate links between rays and samples, the distances between vertices and their directions (Otero et al., 2005). For a full description of compositional biplots and their practical usage, several examples are available in the literature (e.g., Pison et al., 2003; Maronna et al., 2006; Filzmoser and Hron, 2008; Filzmoser et al., 2009a; Hron et al., 2010).

The number of all measured elements (53) was reduced to 18 (Table 1) based on 3 main criteria: 1) the removal of elements with $> 40\%$ of values below the detection limit (DL), 2) choosing arbitrary mostly two representative Rare Earth elements which are geochemically congruent and 3) choosing elements with a communality of extraction higher than 0.5 (50%) or common variances < 0.5 (e.g. Reimann et al., 2002). Based on such criteria, in this paper, we present biplots with the reduced number of variables (18, Table 1) used as inputs for factor analysis.

Factor analysis (FA) was the multivariate statistical tool to explain the correlation structure of the variables through few numbers of factors (Reimann et al., 2002). Furthermore, FA can be successfully used to reveal the elements sources related to their main hypothetical origins. To minimize or eliminate the presence of outliers and spurious correlation (Pawlowsky-Glahn and Buccianti, 2011), isometric logratio transformation (ilr) was applied on raw-data prior to multivariate analysis (Filzmoser et al., 2009b). In order to facilitate the interpretation of results, varimax rotation was implemented, since it is an orthogonal rotation that minimises the number of variables that have high loadings on each factor, simplifying the transformed data matrix and assisting interpretation (Reimann et al., 2002). R-mode factor analysis was performed, and, the different factors obtained were studied and interpreted in accordance with their presumed, i.e. geogenic, anthropogenic or mixed (Tables 2, 3).

GeoDAS™ was used to produce interpolated geochemical maps of the normalised factor scores by means of the multifractal inverse distance weighted (MIDW) algorithm (Cheng et al., 1994; Lima et al., 2003). Multifractality exists in many quantities in the nature sciences from geochemical and geophysical data to brittle structural data where quantities show self-similarity or self-affinity which can be characterised by means of multifractal models (Cheng et al., 1994). MIDW is one of the most widely used methods in the interpolation of geochemical data because it preserves high frequency information, retain local variability taking into consideration both spatial association and local singularity (Cheng et al., 1994; Cheng, 1999; Lima et al., 2003, 2005). Singularity is an index representing the scaling dependency from multifractal point of view, which characterizes how the statistical behaviour of a spatial variable changes as the measuring scale changes (Cheng et al., 1994; Cheng, 1999). Spatial association represents a type of statistical dependency of values at separate locations, and its indexes (e.g. covariance, autocorrelation and semivariogram) have been used to characterize the local structure and variability of surfaces (Cheng et al., 1994). During interpolation and mapping of geochemical variables, both spatial association and scaling are taken into account. More detailed description of the method and their practical usage, one can find several publications (Cheng et al., 1994, 2000; Lima et al., 2003, 2005; Albanese et al., 2007, 2015; Zuo et al., 2015; Zuo and Wang, 2016; Minolfi et al., 2018; Parsa et al., 2017; Rezza et al., 2018). During interpolation we used 5 km search radius and a minimum of 4 points and a map resolution was set to 1 km based on the break point on the nearest distance cumulative plot of sampling sites. Considering that the factor scores values present negative and zero values which are not “log transformable”, a min-max normalisation was applied by scaling the original data within a specified range of features (e.g., ranging from 1 to 100). Min-max normalisation is a linear transformation on the original data without changing their geometrical structure (Han and Kamber, 2001).

The concentration–area (C–A) fractal plot (Cheng et al., 1994, 2000; Cheng, 1999) was used to classify the interpolated factor score maps and capture the different spatial patterns (Figs. 4, 5). Computations (e.g. log-ratio transformations, regressions, and factor analysis) and graphical representations were implemented by the open source statistical software of R and CoDaPack (Comas-Cufí and Thió-Henestrosa, 2011).

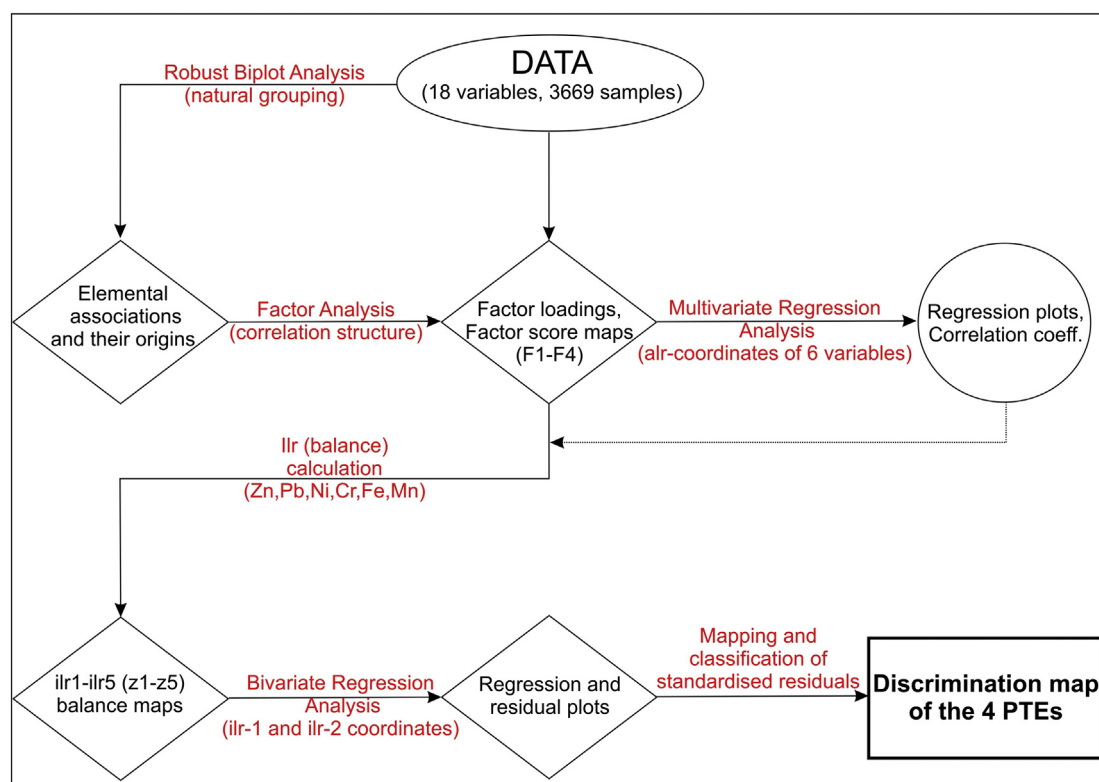


Fig. 2. The flowchart of the compositional discrimination analysis.

2.3. Compositional discrimination analysis

2.3.1. Multivariate regression analysis

Based on elemental associations on robust biplots, multivariate regression analyses were performed using 6 variables (Zn, Pb, Cr, Ni, Fe, and Mn). Iron and Mn were involved as independent variables because both of them and especially their secondary oxy-hydroxides are susceptible of adsorbing Ni, Cr, Zn and Pb in soils (Salminen et al., 2005; Kabata-Pendias, 2010; Reimann et al., 2014). Additive log-ratio (alr) transformation was applied on the 6 variables using Ti as a common divisor, as it is one of the most stable and less mobile elements in soils (Kabata-Pendias, 2010). Additive log-ratio was chosen because it is not

affected by the singularity problem in regression and correlation calculation (Aitchison, 1986; Pawłowsky-Glahn et al., 2015), and it enabled us to reveal the proportional changes for each variable. The alr-transformation is based on the following equation:

$$y = alr(x) = \left[\ln\left(\frac{x_1}{x_D}\right), \ln\left(\frac{x_2}{x_D}\right), \dots, \ln\left(\frac{x_{D-1}}{x_D}\right) \right] \quad (1)$$

where x_1 – x_{D-1} are the $D-1$ variables (all variables except the common divisor element) and x_D is the common divisor element (Aitchison, 1986).

In order to reduce the leveraging effects of outliers on the regression hyperplane, the least trimmed squares (LTS) fitting method was

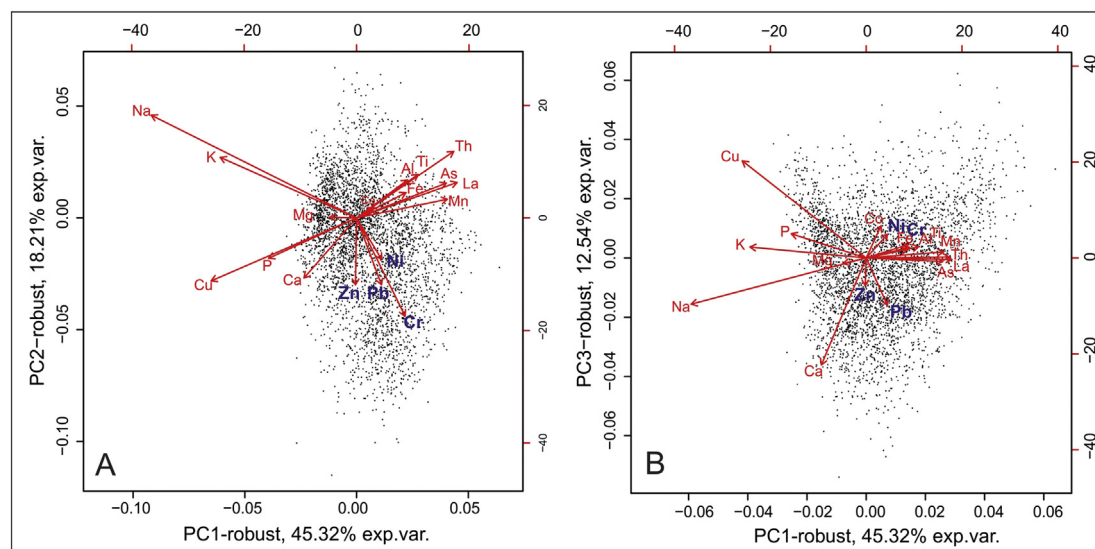


Fig. 3. Robust biplots for the first and second principal components (A) and for the first and third principal components (B) based on the 18 investigated elements.

Table 1

Descriptive statistics of the 18 elements (mg/kg) using 3669 topsoil samples from the Campania Region, Italy. Q1 and Q3 indicate the 1st and 3rd quartiles, respectively.

Variables	Min.	Q1	Mean	Median	Q3	Max.	Std. Deviation	Skewness	Kurtosis
Cu	2.51	32.28	109.32	62.24	128.25	2394.33	156.83	5.49	44.56
Pb	3.12	36.86	73.85	54.17	78.52	2052.18	91.72	7.71	100.16
Zn	11.4	69.70	119.30	91	127.95	3210.60	118.46	9.23	164.68
Ni	0.5	10.00	16.23	14.80	18.30	100.90	10.98	2.42	9.97
Co	0.5	6.90	10.56	10.30	13.50	79	5.02	1.63	12.39
Mn	77	650	863.51	779	972	7975	443.38	5.44	55.78
Fe	1600	19,500	25,031.97	25,100	30,700	154,600	8012.21	1.19	18.26
As	0.6	8.40	12.57	12.10	15.30	163.80	7.35	6.06	87.84
Th	0.3	8.10	12.84	12.40	16.30	60	6.74	1.13	3.47
Ca	800	12,600	35,495.69	22,300	43,500	295,200	36,675.93	2.38	7.15
P	50	725	1641.10	1250	2330	16,620	1250.29	2.20	12.13
La	0.9	31.30	42.02	41.50	51.50	162.3	18.26	0.73	2.46
Cr	0.25	9.20	17.63	14	20.60	808.40	26.16	16.77	386.10
Mg	700	3800	7347.26	5800	8100	104,600	7479.31	5.06	35.80
Ti	5	750	1159.42	1240	1610	3270	618.70	−0.23	−0.65
Al	2100	26,700	40,584.57	41,600	54,050	94,700	17,036.21	0.01	−0.75
Na	20	690	3667.46	2600	5845	29,490	3587.45	1.23	1.68
K	400	4700	14,008.61	9500	18,500	68,200	12,472.37	1.39	1.21

Table 2

Varimax-rotated factors of logratio transformed (ilr) data; bold entries: loading values over absolute value of 0.5.

Variables	Factors				Communalities
	F1	F2	F3	F4	
Cu	−0.12	−0.77	0.23	−0.14	0.67
Pb	−0.28	0.10	0.09	−0.84	0.79
Zn	0.13	−0.21	−0.03	−0.83	0.75
Ni	0.90	−0.13	−0.28	0.09	0.91
Co	0.88	−0.10	0.01	0.37	0.92
Mn	0.69	0.41	0.04	0.05	0.65
Fe	0.79	0.31	0.20	0.27	0.83
As	0.01	0.72	0.25	−0.11	0.60
Th	−0.08	0.72	0.50	0.25	0.84
Ca	0.17	−0.17	−0.87	−0.06	0.83
P	−0.22	−0.72	0.00	0.03	0.57
La	−0.16	0.81	0.37	0.08	0.82
Cr	0.80	−0.08	−0.32	−0.20	0.78
Mg	0.20	−0.03	−0.82	0.14	0.73
Ti	−0.72	0.38	0.35	0.06	0.79
Al	−0.01	0.51	0.53	0.52	0.81
Na	−0.86	−0.24	0.17	0.14	0.84
K	−0.62	−0.46	0.30	0.50	0.94
Eigenvalues	5.81	4.71	2.08	1.31	
Total variance in %	32.26	26.14	11.57	7.30	
Cum. of total variance (%)	32.26	58.40	69.97	77.26	

selected because it is highly robust and a fast algorithm for its computation is available (Rousseeuw and Van Driessen, 2002). In the LTS method the sum of some proportion of the smallest squared residuals is minimised. The proportion may vary from 50 to 100%, with 50% providing maximum protection against outliers. We used the

Table 3

Explanation of the four-factor model extracted from the (ilr) logratio clr back-transformed data, which explains 77.26% of total variance.

Factors	% of variance explained	Association of variables	Interpretation
F1	32.26	(1) Ni, Co, Cr, Fe, Mn (2) Na, Ti, K	1. Pedogenic processes in flysch materials, with adsorption and co-precipitation of Fe and Mn oxy-hydroxides in oxidizing environment (Fig. 3A) 2. Potassic and ultrapotassic rocks and volcano-sedimentary deposits (Fig. 3A)
F2	26.14	(1) La, As, Th, Al (2) Cu, P	1. Pyroclastic rocks of the Mt. Roccamonfina volcano with high levels of incompatible elements (Fig. 3B) 2. Agricultural practices using phosphate fertilisers (Fig. 3B)
F3	11.57	(1) Al, Th (2) Ca, Mg	1. Pyroclastic deposits related to the Mt. Roccamonfina (Fig. 4A) 2. Reflects weathering processes of limestone and dolostone (Fig. 4A)
F4	7.30	(1) Al, K (2) Pb, Zn	1. Potassic pyroclastic composition of the Mt. Somma-Vesuvius (Fig. 4B) 2. Anthropogenic sources like high rate of fossil fuel combustion, industrial and vehicular emission release (Fig. 4B)

compromise of 75% of the data. *t*-Test statistics and its related P-values were calculated at the 95% confidence level to test whether both of the independent variables (Fe/Ti and Mn/Ti) significantly contributed to the regression model or one of them could be dropped. The null hypothesis is that the corresponding model parameter equals 0, based on the extra sums of squares attributable to each variable if it is entered into the model last. F-test statistics and its P values were also calculated to analyse variance and to test whether significant relationship existed between each dependent (Zn/Ti, Pb/Ti, Cr/Ti and Ni/Ti) and independent variable.

2.3.2. Calculation and mapping of balances

Based on the multivariate regression analysis, sequential binary partition was implemented using the same 6 variables (Zn, Pb, Cr, Ni, Fe and Mn) by dividing them into specific groups of non-overlapping elements. Balances are particular ilr-coordinates (isometric-logratio) having orthonormal bases which can be interpreted in the D-1 (D: dimension) real space as ratios of elemental associations (Egozcue et al., 2003). Balances can be calculated using the following formula:

$$z_i = \sqrt{\frac{rs}{r+s}} \ln \left(\frac{\prod_{+} x_j^{1/r}}{\prod_{-} x_k^{1/s}} \right) \text{ for } i = 1, \dots, D-1, \quad (2)$$

where the products \prod_{+} and \prod_{-} only include parts coded with + and −, and r and s are the numbers of positive and negative signs (parts) in the i-th order partition, respectively (Egozcue and Pawłowsky-Glahn, 2005). Table 4 shows the sequential binary partitions of the calculated 5 balances (z1–z5).

The calculated 5 balances (z1–z5) have been normalised by using the linear min-max normalisation method (from 1 to 10), prior to the multifractal interpolation. The interpolated balance maps were classified based on the CA (Concentration-Area) fractal method (Cheng et al.,

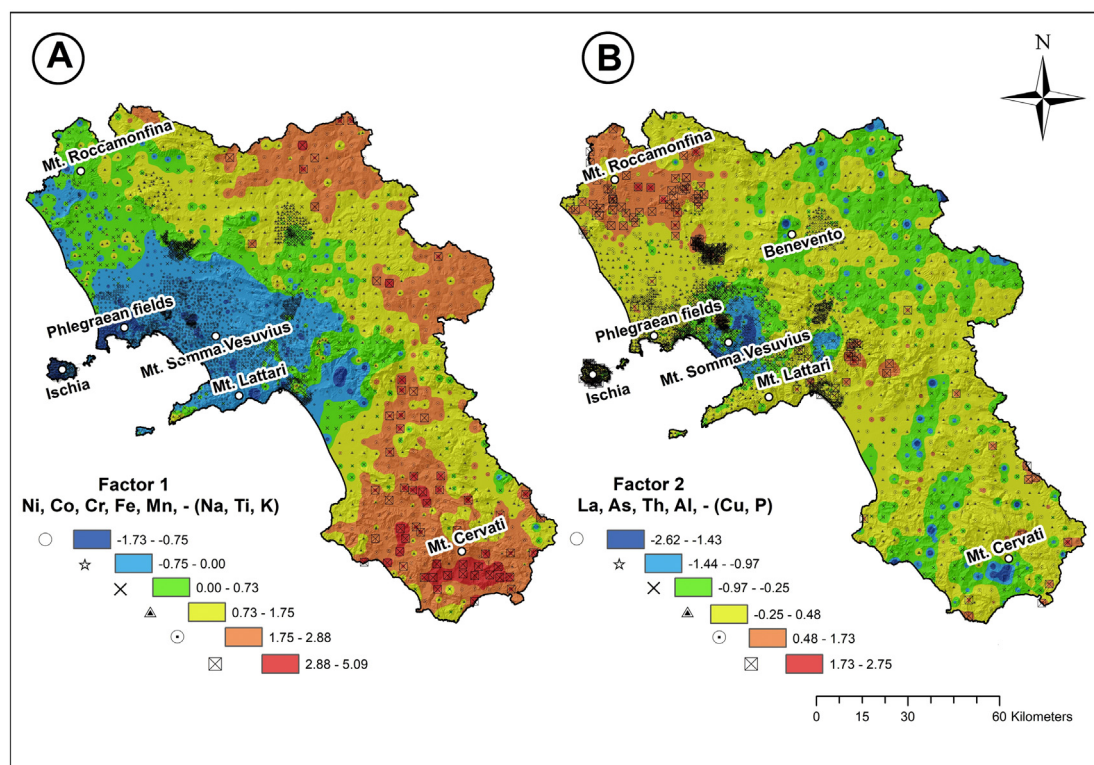


Fig. 4. Interpolated factor score maps of the Factor 1 (A) and Factor 2 (B).

1994; Cheng, 1999) to display the spatial patterns and the proportional changes of elemental associations. In this study we focus and present the first 4 balance maps (z1–z4 or ilr1–ilr4) because these depict the spatial proportion of the investigated PTEs.

2.3.3. Bivariate regression on specific balances and their residual map

Finally, by taking into consideration the factor score and balance maps, and the linear relationship of the 4 PTEs with Fe and Mn, a bivariate regression analysis was performed using normalised ilr-2 (independent) and ilr-1 (dependent) variables. The least trimmed squares

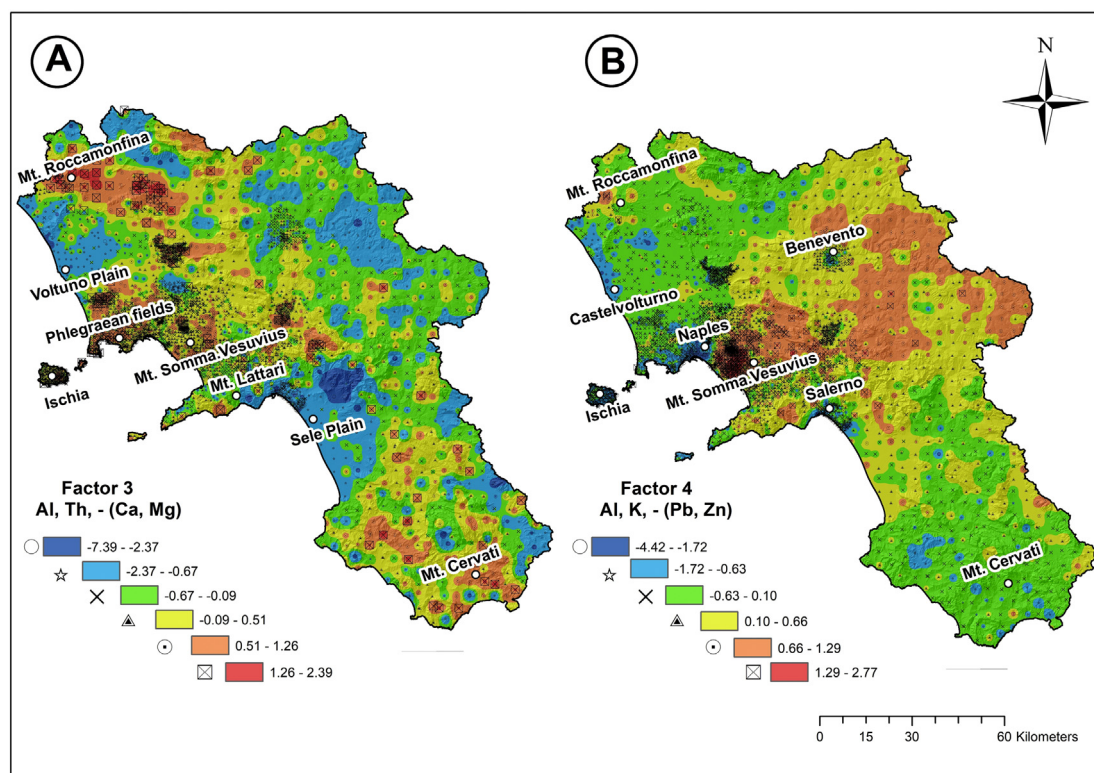


Fig. 5. Interpolated factor score maps of the Factor 3 (A) and Factor 4 (B).

Table 4

Sequential binary partition table of the 6 investigated variables and the obtained balances (z1–z5). Parts coded with + and – are the elemental associations involved in the calculation of the i-th order partition, respectively.

	Zn	Pb	Cr	Ni	Fe	Mn
z1 (ilr-1)	+	+	+	+	–	–
z2 (ilr-2)	+	+	–	–		
z3 (ilr-3)	+	–				
z4 (ilr-4)			+	–		
z5 (ilr-5)					+	–

(LTS) fitting method was also selected here because it is highly robust and resistant against outliers (Rousseeuw and Van Driessen, 2002).

The obtained standardised residuals (residuals divided by standard error) were plotted against ilr-2 coordinates. We were particularly interested in large positive and negative standardised residuals which may indicate different sources of contamination or enrichment. The residuals plot was divided into several parts based on the classification of the ilr-2 and the residuals by means of Tukey's box-and-whiskers plot. The areas below the lower quartile (Q1) and above the upper quartile (Q3) of the standardised variables were called geogenic enrichment and anthropogenic contamination, respectively. These areas were further subdivided based on the lower and upper quartiles of the ilr-2 coordinates which indicate the limits of Cr-Ni and Zn-Pb enrichment or contamination, respectively.

The standardised residuals were interpolated by using the multifractal method and classified by means of the CA-plot (Concentration-Area) to reveal their spatial patterns. The residual map was overlain by those sampling points whose ilr-2 and standardised residual values were beyond or lower than their respective Q1 and Q3. In this way, we could see exactly which elemental associations a contamination or an enrichment is related to.

3. Results and discussion

3.1. Covariance and correlations between variables

The robust biplots explain 63.53% (PC1–PC2) and 57.84% (PC1–PC3) variability of the investigated 18 elements. They reveal several elemental associations related to the underlying geology and anthropogenic activities.

In Fig. 3A, Al-Ti-Th, Fe, Mn, As, La, and Co elemental association (A_1) is characterised by the vicinity of their vertices and their rays point in the same direction. This association is likely to have a mixture of geogenic origin. Within A_1 , Al, Ti and Th vertices are superimposed among which Th has the highest vector length. In fact, A_1 might be divided in two sub-groups, namely A_{1-1} (Al-Ti-Th), which is possibly related to pyroclastic deposits and A_{1-2} (Fe, Mn, Co, As, La), which is probably tied to the coprecipitation of Fe-, Mn-oxy-hydroxides in siliciclastic deposits (Buccianti et al., 2015; Zuzolo et al., 2017).

Sodium and K variables form clearly another association (A_2) because their vertices are located close to each other and they point in the same direction. Both variables display high communalities (the length of the ray). The A_2 may be associated with topsoils underlain by potassic and ultrapotassic rocks.

Copper and P elements belong to the A_3 because their vertices overlap each other. Furthermore, A_2 and A_3 associations are geometrically symmetric with respect to the Mg vector meaning that A_3 enhances a mixed behaviour of Cu and P which is likely related to volcanic materials and phosphate fertilisers.

Calcium and Mg vectors are symmetric around the A_3 and form a separate group of A_4 . In fact, these two elements are dominants in carbonate rocks.

Zinc, Pb, Ni and Cr variables form another association (A_5); their vertices lay closer to each other and point to the same direction. They

may be related to anthropogenic sources like vehicular emission and industrial activities. However, Fig. 3B clearly shows the discrimination of the A_5 , where Pb and Zn vertices remain close to each other, but Ni and Cr changed their position and joined the elemental association of the A_1 . This could prove the affinity of Ni-Cr to the coprecipitation of Fe-, Mn-oxy-hydroxides in siliciclastic material.

Compositional factor analysis was used to discern the correlations and interrelationships between variables, and their possible geochemical source patterns in the Campanian topsoils.

The total variance of the 18 elements is 77.26% in the four-factor model, which was chosen based on the break-point on the scree-plot of all factors. The 4 factors, named F_1 , F_2 , F_3 and F_4 , account for 32.26%, 26.14%, 11.57% and 7.30% variability, respectively (Table 2). Variables with loadings over the absolute value of 0.5 have been considered to describe the main composition of each factor. All variables hold communalities over 0.5 (50% of variability) meaning that the 4 factor models capture quite well the elemental interrelationships and their possible geogenic and/or anthropogenic sources (Table 3). The 18 elements of the four-factor model were separated by positive and negative loadings and sorted in descending order:

F_1 : Ni, Co, Cr, Fe, Mn, - (Na, Ti, K)

F_2 : La, As, Th, Al, - (Cu, P)

F_3 : Al, Th, - (Ca, Mg)

F_4 : Al, K, - (Pb, Zn)

The Fig. 4A shows interpolated map of factor scores (F_1), ranging from –1.73 to 5.09. High factor scores (ranging from 2.88 to 5.09) are mapped in the eastern and south-western parts of the study area where siliciclastic deposits prevail (Figs. 1, 4A). Pedogenic processes in flysch materials (fine-size) are mostly characterised by adsorption and coprecipitation of Co, Cr and Ni with Fe and Mn oxy-hydroxides in oxidizing environment (Postma, 1985). Low factor scores (ranging from –1.73 to –0.75) are revealed close to large volcanic complexes like the Phlegraean Fields, Mt. Somma-Vesuvius and Mt. Roccamonfina. Negative scores are given by the antithetic elemental association including Ti, Na, K. These elements are likely related to the potassic and ultrapotassic rocks and volcano-sedimentary deposits from major sector collapse of volcanoes in the survey area (Scheib et al., 2014; De Vivo et al., 2016).

High factor scores of the F_2 range from 1.73 to 2.75 and are related to the positive correlation between Al, La, As, and Th in soils of the Mt. Roccamonfina and surroundings (Fig. 4B). Pyroclastic rocks of the Mt. Roccamonfina volcano are characterised by high levels of incompatible trace elements with an orogenic signature having troughs at Ba, Ta, Nb, and Ti, and peaks at Cs, K, Th, U, and Pb (Conticelli et al., 2009). Low factor scores (ranging from –2.62 to –1.43) can be observed around the Mt. Somma-Vesuvius, Benevento (Vitaliano municipality), and in the Mt. Cervati. Large agricultural fields like vineyards and orchards are located in these areas. Copper and P are likely related to agricultural practices using phosphate fertilisers (Thiombane et al., 2017).

The F_3 factor score map (Fig. 5A) shows (particularly for Al and Th elemental association) elevated values (ranging from 1.26 to 2.39) close to the Mt. Roccamonfina. The Mt. Roccamonfina volcano is characterised by pyroclastic rocks with high level of trace elements such Th, U, and La (Conticelli et al., 2009). But relatively high factor scores (ranging from 0.51 to 1.26) can be observed around the Phlegraean fields and the Mt. Somma-Vesuvius, and along the Apennines which were related to the occurrences of Campanian Ignimbrites (De Vivo et al., 2010). In contrast, low factor scores (Ca and Mg elements) are mapped exactly in correspondence with the three large river plains of the Campania Region, namely Volturno, Sele and Telesina. Furthermore, these areas are well drained by rivers and tributaries which spring from the carbonate rocks of the Apennine chain. Indeed, Ca and Mg in these areas are mostly associated with the weathering processes of limestone and dolostone (Peccerillo, 2005; Albanese et al., 2007).

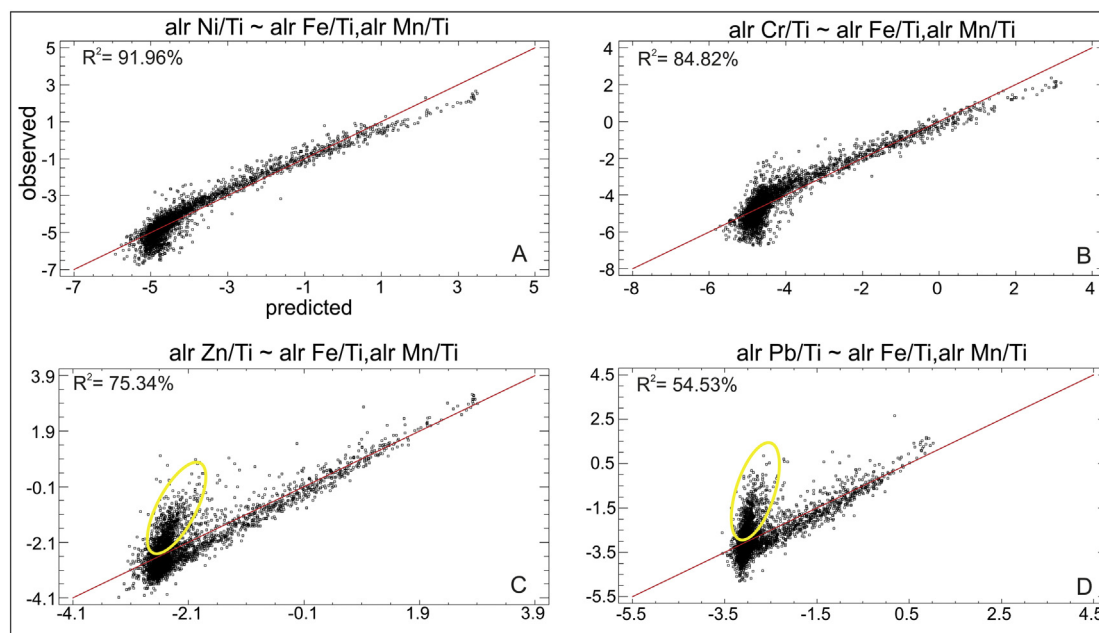


Fig. 6. Multivariate regression analysis of the alr-transformed Ni, Cr, Zn and Pb (dependent variable) against Fe, Mn (independent variables). Note the good correspondence and strong linear relationships between predicted and observed values. The high positive correlation (R^2) decreases from Ni to Pb. In addition, the yellow ellipses at Zn and Pb mark deviation tails from linearity suggesting different processes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The F_4 score map (Fig. 5B) shows high positive values (ranging from 1.29 to 2.77) around the Mt. Somma-Vesuvius, where Al and K are the most congruent elements with the potassic pyroclastics composition (Peccerillo, 2001; De Vivo et al., 2010). Low factor scores (particularly Pb and Zn elemental association) are mapped exclusively on topsoils of the most urbanised areas (Naples, Salerno and Castelvoturno) of the study area. Lead and Zn may be tied to anthropogenic sources such high rate of fossil fuel combustion, industrial and vehicular emissions release.

3.2. Compositional discrimination analysis

In this session, we present the results of compositional data analysis used to discriminate the main sources of the investigated potentially toxic elements (Zn, Pb, Cr and Ni).

3.2.1. Multivariate regression analysis

Multivariate regression analysis was implemented by using the alr-transformed coordinates of 6 elements (Zn, Pb, Cr, Ni, Fe and Mn) in order to examine their relationships and correlations (Fig. 6). The multivariate regression analysis revealed strong positive correlation ($R^2 > 54.53\%$) and statistically significant (F-test statistics, at 95% confidence level) linear relationship between each dependent (Zn/Ti, Pb/Ti, Cr/Ti, Ni/Ti) and independent (Fe/Ti, Mn/Ti) alr-coordinates (Fig. 6). Based on the t -test statistics and its associated P-values at the 95% confidence level, both of the independent variables significantly contribute to the predicted model results, hence neither of them can be dropped.

It should be noticed that the positive correlation decreased continuously from Ni/Ti (R^2 : 91.96%), Cr/Ti (84.82%) to Zn/Ti (75.34%) and Pb/Ti (54.53%) (Fig. 6). In addition, interesting deviation tails from the linear trends were discerned in case of Zn/Ti and Pb/Ti which suggest different processes (Figs. 6C, D). The deviations can be observed at low alr-Fe/Ti and alr-Mn/Ti which may indicate other sources (anthropogenic?) of Zn and Pb besides the accumulation of Fe-, and Mn-oxy-hydroxides.

3.2.2. Balance (ilr) maps

Based on the multivariate regression analysis, the same 6 elements (Zn, Pb, Cr, Ni, Fe and Mn) have been chosen to perform sequential binary partition and obtain balances (specific ilr-coordinates) (Table 4).

3.2.3. Ilr-1 map (ZnPbCrNi/FeMn)

The first balance (ilr1) map reveals higher proportion of the PTEs around large urban and industrial areas (e.g. Naples, Salerno, and Castelvoturno). These cities are located all along the coastline where the highest population density and the majority of industrial settlements and traffic are concentrated (Fig. 7A).

In contrast, the higher abundance of Fe and Mn can be observed close to large volcanic edifices (e.g. Mt. Roccamonfina, Mt. Somma-Vesuvius) and their surroundings. Some high Fe and Mn proportion patches are also displayed in the eastern siliciclastic zone indicating that the behaviour and accumulation of these elements are mainly influenced by the underlying bedrocks (Fig. 7A).

3.2.4. Ilr-2 map (ZnPb/CrNi)

The second balance map is based on the subdivision of PTEs taking into consideration their separate associations on the robust biplot (PC1–PC3) and their factor loadings (Fig. 3, Table 2). The higher proportion of Zn and Pb is associated with urban and industrial areas (e.g. Naples and Salerno) but also reaches higher abundance close to large volcanic complexes like Phlegraean Fields and Ischia (Fig. 7B).

In contrast, the higher proportion of Cr and Ni can be observed exclusively in the eastern and southern part of the study area, far away from most urbanised areas (Fig. 7B). These elements have higher abundance in topsoils over siliciclastic deposits (e.g. flysch sediments) in the Apennine Mts; hence they are much related to specific lithology (e.g. flysch, claystone) and related pedogenic processes (Figs. 1A, 7B). According to Bucciante et al. (2015), the Ni and Cr abundance in Campania is highly influenced by dispersion mechanisms that characterize sediment material particularly flysch deposits.

It is interesting to see the NW-SE trending edge in the middle of the study area which separates the two different proportion zones and coincides roughly with the boundary of siliciclastic and carbonate

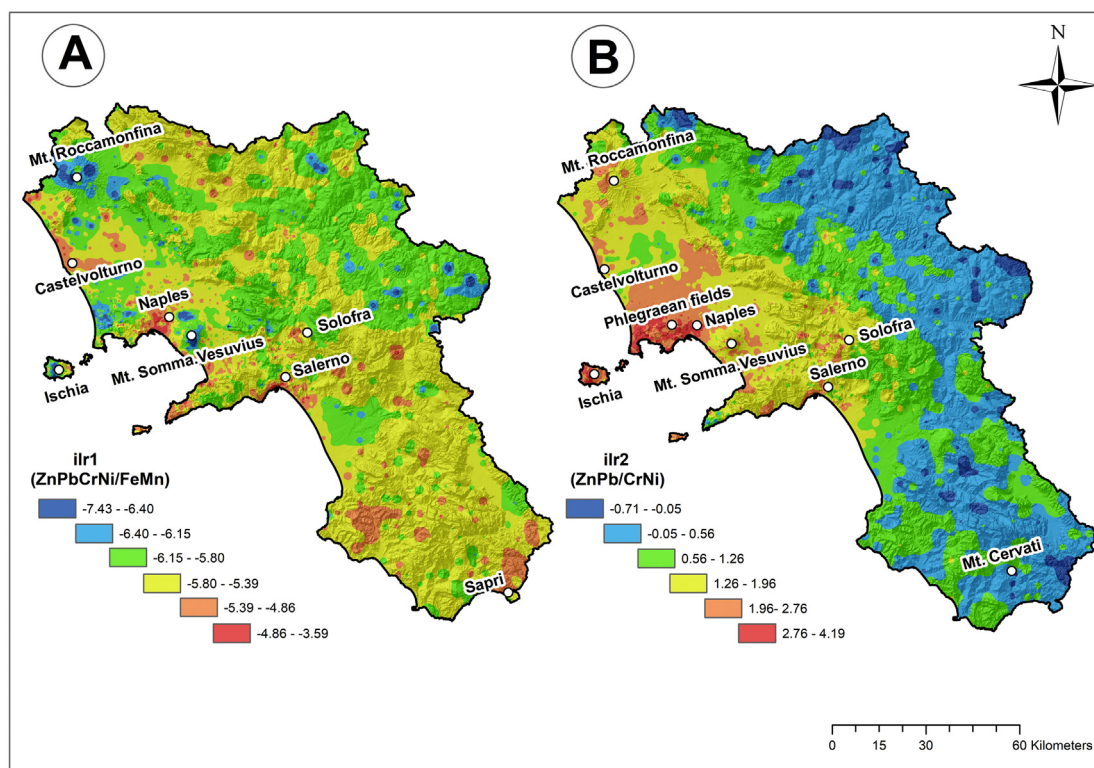


Fig. 7. (A) The interpolated ilr-1 map. Note the higher proportion of the PTEs (with red colour) around Naples and Salerno. (B) The interpolated ilr-2 map. The higher spatial abundance of Cr and Ni (with blue colour) is clearly related to the eastern and southern siliciclastic zones. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

outcrops (Figs. 1A, 7B).

3.2.5. Ilr-3 map (Zn/Pb)

The third balance map is based on the ratio between Zn and Pb and sheds light on their spatial abundance (Fig. 8A). The map unveils that the higher Zn proportion zones are mainly related to underlying siliciclastic deposits in the eastern and southern part of the study area, far away from urbanised regions (Fig. 8A). According to Minolfi et al. (2018), the behaviour and accumulation of Zn is mainly influenced by the underlying bedrocks in the Campania Region. Buccianti et al. (2015) revealed that Zn has one the lowest clr-variances (central log-ratio) meaning that it maintains similar abundance in the Campanian topsoils.

The higher Pb proportion zones are located in the urbanised area of Naples and some parts of Ischia and Sorrento Peninsula where they are tied to intensive traffic (Fig. 8A). The highest concentration of Pb (up to 425 mg/kg) in stream sediment was also described from Naples and that was explained by heavy traffic and human activities (Lima et al., 2003; De Vivo et al., 2016; Minolfi et al., 2018). However, the higher Pb proportion zone can also be revealed around Mt. Roccamonfina, where the average Pb concentration from phonolite and phonotheprite was 129 mg/kg (Peccerillo, 2005). This concentration is two times higher than the counterpart of Campanian volcanic rocks (64 mg/kg) (Paone et al., 2001); hence the Pb enrichment in the Roccamonfina area is more likely related to the underlying volcanic rocks.

3.2.6. Ilr-4 map (Cr/Ni)

The fourth balance map demonstrates the relative abundance of Cr and Ni and their spatial distribution (Fig. 8B). The map shows that their proportion is similar in the vast majority of the study area (ilr-4: -0.2 – 0.25). One of the higher Cr proportion areas is related to the Sarno Basin (e.g. Solofra district) where several factories (e.g. tannery industry) operate contributing to the abundance of Cr (Fig. 8B). Some

high Cr proportion patches are observed around the Mt. Roccamonfina and above some carbonate massifs (e.g. Mt. Matese) where the most advanced pedogenetic processes (e.g. rubification, eluviation) were demonstrated (Petrik et al., 2018). The trivalent chromium (Cr^{3+}) behaves much like Al^{3+} and Fe^{3+} and tends to accumulate with secondary oxy-hydroxides in matured soils especially under alkaline environment (Salminen et al., 2005; Reimann et al., 2014).

Nickel reaches higher proportion compared to Cr in the eastern part of the study area in topsoils over claystone and siltstone (Figs. 1A, 8B). The lowest mobility and variance of Ni concentrations were described here by Petrik et al. (2017). The higher abundance of Ni close to the Mt. Somma-Vesuvius may be related to underlying volcanic rocks.

3.2.7. Bivariate regression analysis and residual map

Based on the factor score and balance maps and the strong positive correlation (R^2 : 54.53–91.96%) of the 4 PTEs with Fe and Mn, we performed bivariate regression analysis using the normalised ilr-2 (independent) and ilr-1 (dependent) variables to discriminate and map the anthropogenic and geogenic parts of the contamination (Fig. 9). We were particularly interested in standardised residuals which were plotted against normalised ilr-2 coordinates (Fig. 9B). The large positive (> 1.55) standardised residuals reveal low FeMn proportion (ilr1), hence mainly indicating anthropogenic contamination. In contrast, large negative (< -1.04) standardised residuals (high FeMn proportion) represent geogenic contamination (Fig. 9B).

The standardised residuals of the bivariate regression were used to make residual map which was classified by means of the C-A (Concentration-Area) plot (Fig. 10). The residual map was overlain by those sampling points which are related to different sources of contamination or enrichment based on the standardised residuals plot (Figs. 9B, 10). The large positive residuals (> 1.55 , with red colour) all indicate anthropogenic contamination of the 4 PTEs (Figs. 9B, 10). The densely populated, large urban and industrial areas along the coastline

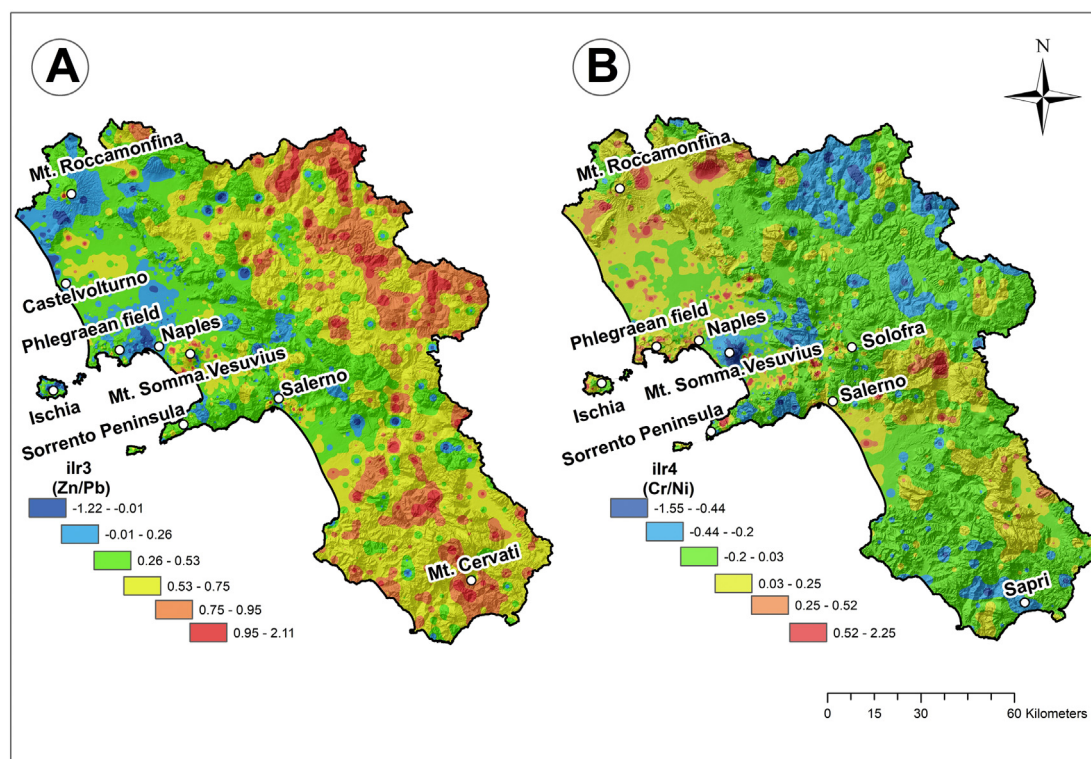


Fig. 8. (A) The interpolated ilr-3 map. Note the higher proportion of Zn (with red colour) in the eastern and southern siliciclastic zones. Pb proportion is the highest around Naples. (B) The interpolated ilr-4 map. The higher spatial abundance of Ni is related to the Mt. Somma Vesuvius and the eastern siliciclastic zone. The higher Cr proportion can be observed around Solofra, for example. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(e.g. Naples, Salerno and Castelvolturno) are all displayed as highly contaminated regions mainly due to the heavy traffic (Pb anomaly) and various heavy industries (e.g. production of alloys: Zn). Some scattered, high contamination patches in the Sarno Basin are likely related to intensive agriculture using phosphate fertilisers (Zn, Cr, Ni anomaly). Contamination in the Solofra industrial district is linked to traffic emission (Pb), alloy production (Zn and Ni) and tannery (Cr and Ni) industries (Fig. 10). The southern urbanised region around Sapri is also displayed as contaminated area, which was associated with large urban areas (Pb) and smelting industries (Cr, Ni) (Thiombane et al., 2017).

In contrast, the large negative residuals (< -1.04 with dark blue) indicate the geogenic source of the four investigated PTEs. The large volcanic complexes (e.g. Mt. Somma-Vesuvius, Phlegraean Fields, Mt. Roccamonfina) are all characterised by negative residuals demonstrating the main influence of underlying geology (e.g. different

volcanic rocks) on the proportion of Zn and Pb (Fig. 10). Some negative residuals can be discerned in the eastern part of the study area where the anomaly of Cr and Ni may be influenced not only by the presence of Fe-, and Mn-oxy-hydroxide but also the organic matter and clay content (Fig. 10). According to Buccianti et al. (2015), the Ni and Cr distribution are strongly dependent on the various lithology (e.g. clay content) and grain size of the siliciclastic zone in the Apennine Mts.

4. Conclusion

This study demonstrates a comprehensive discrimination analysis carried out on 4 PTEs (Zn, Pb, Cr and Ni) of the Campanian high-density topsoil data set. A workflow of compositional data analysis was implemented ranging from robust biplot and factor score maps to multivariate and bivariate regression analyses to discriminate the possible

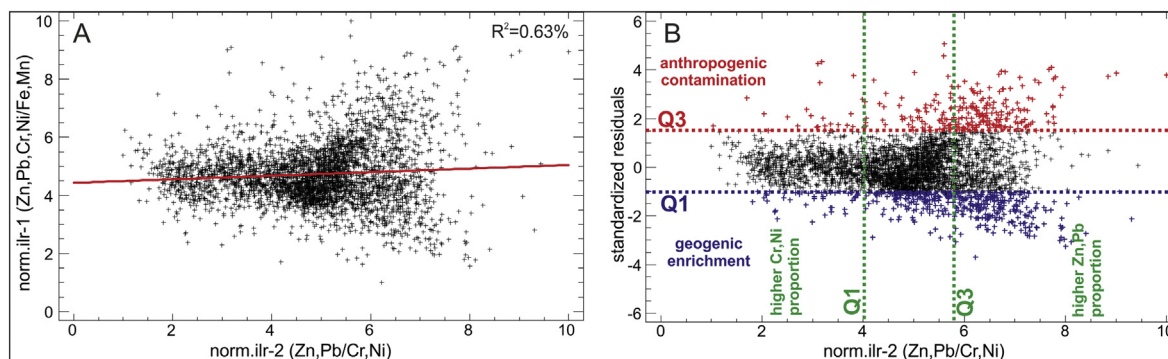


Fig. 9. (A) Bivariate regression analysis between normalised ilr-2 (independent) and ilr-1 (dependent) variables. (B) Standardised residuals against normalised ilr-2. The plot was divided into more parts based on the classification of ilr-2 and residuals by means of Tukey's box-and-whiskers plot. Areas outside of lower (Q1) and upper quartiles (Q3) belong to geogenic enrichment and anthropogenic contamination of different elemental associations.

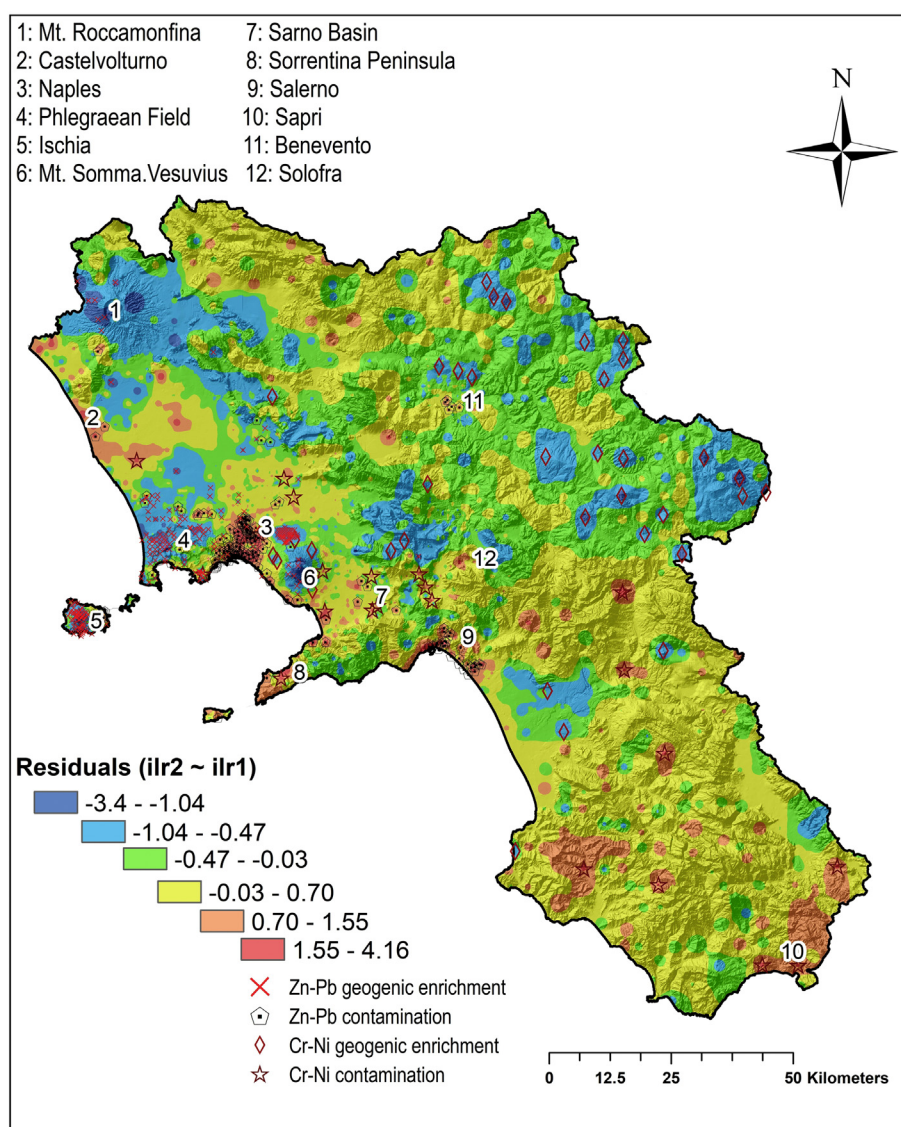


Fig. 10. Classified standardised residual map indicating the contamination and enrichment areas. The map is overlain by those sampling points which are related to anthropogenic contamination or geogenic enrichment of the respective 4 PTEs.

sources of contamination/enrichment.

The study represents a new discrimination index based on standardised residuals of the bivariate regression of the ilr-2 against ilr-1 coordinates which enabled us to identify and separate the anthropogenic contamination from the geogenic enrichment of the respective 4 PTEs. The large urban and industrial areas (e.g. Naples, Salerno) along the coastline are mainly contaminated by Pb and Zn due to heavy traffic and alloy production. Some Cr and Ni contamination was discerned in the Sarno Basin due to the release of waste rich in Cr-Ni deriving from tannery industries (e.g. Solofra).

The large volcanic complexes (e.g. Mt. Somma-Vesuvius, Mt. Roccamonfina) are all characterised by geogenic enrichment of Zn and Pb. In contrast, Cr and Ni-enrichment is mainly related to the siliciclastic deposits where their proportion is not only influenced by Fe and Mn but also organic matter, clay content and dispersion mechanism.

Acknowledgements

This work was supported by the financial support of the MIUR through two industrial projects: 1) “Integrated agro-industrial chains with high energy efficiency for the development of eco-compatible processes of

energy and biochemicals production from renewable sources and for the land valorization (EnerbioChem)” PON01_01966, funded in the framework of Operative National Program Research and Competitiveness 2007–2013 D. D. Prot. n. 01/Ric. 18.1.2010; and 2) “Development of green technologies for production of BIOchemicals and their use in preparation and industrial application of POLlmeric materials from agricultural biomasses cultivated in a sustainable way in Campania Region (BioPolis)” PON03PE_00107_1, funded in the framework of Operative National Program Research and Competitiveness 2007–2013 D. D. Prot. N. 713/Ric. 29/10/2010 (Research Units Responsible: Prof. B. De Vivo). The work has been also supported by the Istituto Zooprofilattico Sperimentale del Mezzogiorno (IZSM) by means of the Contract with the Centro Interdipartimentale di Ricerca Ambientale (CIRAM) in the framework of the “Campania Trasparente - Attività di monitoraggio integrato per la Regione Campania” project funded in the framework of Del.G.R. n. 497/2013: Fondo per le Misure Anticicliche e la Salvaguardia dell'occupazione - Azione B4 “Mappatura del Territorio” approved with Executive Decree DG “Sviluppo Economico” n.585, 14/09/2015 (Research Unit Responsible: Prof. B. De Vivo).

References

- Adamo, P., Zampella, M., 2008. Chemical speciation to assess potentially toxic metals (PTMs) bioavailability and geochemical forms in polluted soils. In: De Vivo, B., Belkin, H.E., Lima, A. (Eds.), *Environmental Geochemistry: Site Characterization, Data Analysis and Case Histories*. Elsevier, pp. 175–212.
- Agterberg, F.P., 2001. Multifractal simulation of geochemical map patterns. In: Merriam, D.F., Davis, J.C. (Eds.), *Geologic Modeling and Simulation. Computer Applications in the Earth Sciences*. Springer, Boston, MA, pp. 327–346. http://dx.doi.org/10.1007/978-1-4615-1359-9_17.
- Aitchison, J., 1986. *The Statistical Analysis of Compositional Data*. Chapman & Hall, London, pp. 416.
- Albanese, S., De Vivo, B., Lima, A., Cicchella, D., 2007. Geochemical background and baseline values of toxic elements in stream sediments of Campania region (Italy). *J. Geochem. Explor.* 93 (1), 21–34.
- Albanese, S., Sadeghi, M., Lima, A., Cicchella, D., Dinelli, E., Valera, P., Falconi, M., Demetriades, A., De Vivo, B., The GEMAS Project Team, 2015. GEMAS: cobalt, Cr and Ni distribution in agricultural and grazing land soil of Europe. *J. Geochem. Explor.* 154, 81–93.
- Bollen, K.A., 1987. Outliers and improper solutions: a confirmatory factor analysis example. *Sociol. Methods Res.* 15, 375–384.
- Bonardi, G., D'Argenio, D., Perrone, V., 1998. Carta geologica dell'Appennino meridionale. *Mem. Soc. Geol. Ital.* 41.
- Buccianti, A., Lima, A., Albanese, S., Cannatelli, C., Esposito, R., De Vivo, B., 2015. Exploring topsoil geochemistry from the CoDA (compositional data analysis) perspective: the multi-element data archive of the Campania Region (Southern Italy). *J. Geochem. Explor.* 159, 302–316.
- Bureau Veritas Minerals, 2017. AQ250 (Ultra Trace Geochemical Aqua Regia Digestion). pp. 1–2.
- Cheng, Q., 1999. Multifractality and spatial statistics. *Comput. Geosci.* 25, 949–961.
- Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. *J. Geochem. Explor.* 51, 109–130.
- Cheng, Q., Xu, Y., Grunsky, E., 2000. Integrated spatial and spectrum method for geochemical anomaly separation. *Nat. Resour. Res.* 9, 43–56.
- Chester, R., Stoner, J.H., 1973. Pb in particulates from the lower atmosphere of the eastern Atlantic. *Nature* 245 (27–2), 8.
- Cicchella, D., De Vivo, B., Lima, A., 2005. Background and baseline concentration values of elements harmful to human health in the volcanic soils of the metropolitan and provincial area of Napoli (Italy). *Geochem. Explor. Environ. Anal.* 5, 29–40.
- Comas-Cufí, M., Thió-Henestrosa, S., 2011. CoDaPack 2.0: a stand-alone, multi-platform compositional software. In: Egozcue, J.J., Tolosana-Delgado, R., Ortego, M.I. (Eds.), *CoDaWork'11: 4th International Workshop on Compositional Data Analysis*. SantFeliu de Guíxols.
- Conticelli, S., Marchionni, S., Rosa, D., Giordano, G., Boari, E., Avanzinelli, R., 2009. Shoshonite and sub-alkaline magmas from an ultrapotassic volcano: Sr–Nd–Pb isotope data on the Roccamonfina volcanic rocks, Roman Magmatic Province, Southern Italy. *Contrib. Mineral. Petrol.* 157 (1), 41–63.
- De Vivo, B., Petrosino, P., Lima, A., Rolandi, G., Belkin, H.E., 2010. Research progress in volcanology in Neapolitan area, Southern Italy: a review and alternative views. *Mineral. Petrol.* 99, 1–28.
- De Vivo, B., Lima, A., Albanese, S., Cicchella, D., Rezza, C., Civitillo, D., Minolfi, G., Zuzolo, D., 2016. Atlante geochemico-ambientale dei suoli della Campania (Environmental Geochemical Atlas of Campania Soils). Aracne Editrice, Roma, pp. 364 (ISBN 978-88-548-9744-1; [in Italian]).
- Ducci, D., Tranfaglia, G., 2005. L'impatto dei cambiamenti climatici sulle risorse idriche sotterranee in Campania. *Boll. Ordine Geologi Campania* 1–4, 13–21 (in Italian).
- Egozcue, J.J., Pawłowsky-Glahn, V., 2005. Groups of parts and their balances in compositional data analysis. *Math. Geol.* 37 (7), 795–828.
- Egozcue, J.J., Pawłowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. *Math. Geol.* 35 (3), 279–300.
- Fabian, C., Reimann, C., Fabian, K., Birke, M., Baritz, R., Haslinger, E., The GEMAS Project Team, 2014. GEMAS: spatial distribution of the pH European agricultural and grazing land soil. *Appl. Geochem.* 48, 207–216. <http://dx.doi.org/10.1016/j.apgeochem.2014.07.017>.
- Filzmoser, P., Hron, K., 2008. Outlier detection for compositional data using robust methods. *Math. Geosci.* 40 (3), 233–248.
- Filzmoser, P., Hron, K., Reimann, C., 2009a. Principal component analysis for compositional data with outliers. *Environmetrics* 20 (6), 621–632.
- Filzmoser, P., Hron, K., Reimann, C., 2009b. Univariate statistical analysis of environmental (compositional) data - problems and possibilities. *Sci. Total Environ.* 407, 6100–6108.
- Filzmoser, P., Hron, K., Tolosana-Delgado, R., 2012. Interpretation of multivariate outliers for compositional data. *Comput. Geosci.* 39, 77–85.
- Gabriel, K.R., 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58 (3), 453–467.
- Han, J., Kamber, M., 2001. *Data Mining: Concepts and Techniques*. Morgan-Kaufmann Academic Press, San Francisco, pp. 740.
- Hron, K., Templ, M., Filzmoser, P., 2010. Imputation of missing values for compositional data using classical and robust methods. *Comput. Stat. Data Anal.* 54 (12), 3095–3107.
- ISTAT, 2016. Resident population on 1st January, 2017. <https://www.istat.it/en/population-and-households>.
- Kabata-Pendias, A., 2010. *Trace Elements in Soils and Plants*, 4th edition. CRC Press, pp. 548.
- Kürzl, H., 1988. Exploratory data analysis: recent advances for the interpretation of geochemical data. *J. Geochem. Explor.* 30, 309–322.
- Lima, A., De Vivo, B., Cicchella, D., Cortini, M., Albanese, S., 2003. Multifractal IDW interpolation and fractal filtering method in environmental studies: an application on regional stream sediments of Campania Region (Italy). *Appl. Geochem.* 18 (12), 1853–1865. [http://dx.doi.org/10.1016/S08832927\(03\)00083-0](http://dx.doi.org/10.1016/S08832927(03)00083-0).
- Lima, A., Albanese, S., De Vivo, B., 2005. Geochemical baselines for the radioelements K, U and Th in the Campania region, Italy: a comparison of stream-sediment geochemistry and gamma-ray surveys. *Appl. Geochem.* 20, 611–625.
- Maronna, R., Martin, R., Yohai, V., 2006. *Robust Statistics: Theory and Methods*. John Wiley, 978-0-470-01092-1pp. 436.
- Minolfi, G., Albanese, S., Lima, A., Tarvainen, T., Fortelli, A., De Vivo, B., 2016. A regional approach to the environmental risk assessment - human health risk assessment: case study in the Campania region. *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.jgexplo.2016.12.010>.
- Minolfi, G., Petrik, A., Albanese, S., Lima, A., Rezza, C., De Vivo, B., 2018. Lead, Cu and Zn distributions in topsoils of the Campania Region, Italy. In: *Special Issue of the Geochemistry: Exploration, Environment and Analysis*, <http://dx.doi.org/10.1144/geochem.2017-074>.
- Müller, G., 1979. Schwermetalle in den Sedimenten des Rheins—Veränderungen seit 1971. *Umschau* 24, 773–778 (German).
- Otero, N., Tolosana-Delgado, R., Solera, A., Pawłowsky-Glahn, V., Canals, A., 2005. Relative vs. absolute statistical analysis of compositions: a comparative study of surface waters of a Mediterranean river. *Water Res.* 39, 1404–1414.
- Paone, A., Ayuso, R.A., De Vivo, B., 2001. A metallogenic survey of alkalic rocks of Mt. Somma-Vesuvius volcano. *Mineral. Petrol.* 73, 201–233.
- Parsa, M., Maghsoudi, A., Yousefi, M., Carranza, J.M., 2017. Multifractal interpolation and spectrum-area fractal modeling of stream sediment geochemical data: implications for mapping exploration targets. *J. Afr. Earth Sci.* 128, 5–15.
- Pawłowsky-Glahn, V., Buccianti, A., 2011. *Compositional Data Analysis: Theory and Applications*. John Wiley & Sons, pp. 400.
- Pawłowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R., 2015. *Modelling and Analysis of Compositional Data*. 4. John Wiley & Sons, pp. 252.
- Peccherillo, A., 2001. Geochemical similarities between the Vesuvius, Phlegraean fields and Stromboli volcanoes: petrogenetic, geodynamic and volcanological implications. *Mineral. Petrol.* 73, 93. <http://dx.doi.org/10.1007/s007100170012>.
- Peccherillo, A., 2005. Plio-quaternary volcanism in Italy. In: *Petrology, Geochemistry, Geodynamics*. Springer-Verlag, Berlin Heidelberg, 978-3-540-29092-6, pp. 365.
- Petrik, A., Jordan, Gy., Albanese, S., Lima, A., Rolandi, R., De Vivo, B., 2017. Spatial pattern analysis of Ni concentration in topsoils in the Campania Region (Italy). *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.jgexplo.2017.09.009>.
- Petrik, A., Albanese, S., Lima, A., De Vivo, B., 2018. The spatial pattern of beryllium and its possible origin using compositional data analysis on a high-density topsoil data set from the Campania Region (Italy). *Appl. Geochem.* <http://dx.doi.org/10.1016/j.apgeochem.2018.02.008>.
- Pison, G., Rousseeuw, P.J., Filzmoser, P., Croux, C., 2003. Robust factor analysis. *J. Multivar. Anal.* 84, 145–172.
- Postma, D., 1985. Concentration of Mn and separation from Fe in sediments—I. Kinetics and stoichiometry of the reaction between birnessite and dissolved Fe(II) at 10 °C. *Geochim. Cosmochim. Acta* 49 (4), 1023–1033. [http://dx.doi.org/10.1016/0016-7037\(85\)90316-3](http://dx.doi.org/10.1016/0016-7037(85)90316-3).
- Reimann, C., de Caritat, P., 2005. Distinguishing between natural and anthropogenic sources for elements in the environment: regional geochemical surveys versus enrichment factors. *Sci. Total Environ.* 337, 91–107.
- Reimann, C., Filzmoser, P., Garrett, R., 2002. Factor analysis applied to regional geochemical data: problems and possibilities. *Appl. Geochem.* 17 (3), 185–206.
- Reimann, C., Filzmoser, P., Garrett, R.G., Dutter, R., 2008. *Statistical data analysis explained. In: Applied Environmental Statistics With R*. Wiley, Chichester, 978-0-470-98581-6, pp. 362.
- Chemistry of Europe's agricultural soils — part A: methodology and interpretation of the GEMAS data set. In: Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., O'Connor, P., GEMAS Team (Eds.), *Geologisches Jahrbuch (Reihe B)*, Schweizerbart: Hannover, pp. 528.
- Rezza, C., Petrik, A., Albanese, S., Lima, A., Minolfi, G., De Vivo, B., 2018. Mo, Sn and W patterns in topsoils of the Campania Region, Italy. In: *Special Issue of the Geochemistry: Exploration, Environment and Analysis*, <http://dx.doi.org/10.1144/geochem.2017-061>.
- Roland, G., Bellucci, F., Heizler, M.T., Belkin, H.E., De Vivo, B., 2003. Tectonic controls on genesis of ignimbrites from the Campanian Volcanic Zone, Southern Italy. In: De Vivo, B., Scandone, R. (Eds.), *Ignimbrites of the Campania Plain, Italy*. Mineral. Petrol. 79, pp. 3–31.
- Rousseeuw, P.J., Van Zomeren, B.C., 1990. Unmasking multivariate outliers and leverage points. *J. Am. Stat. Assoc.* 85 (411), 633–639. <http://dx.doi.org/10.1080/01621459.1990.10474920>.
- Rousseeuw, P.J., Van Driessen, K., 2002. Computing LTS regression for large data sets. *Statistica* 54, 163–190.
- Salminen, R., Batista, M.J., Bidovec, M., Demetriades, A., De Vivo, B., De Vos, W., Duris, M., Gilucis, A., Gregorauskiene, V., Halamic, J., Heitzmann, P., Lima, A., Jordan, G., Klaver, G., Klein, P., Lis, J., Locutura, J., Marsina, K., Mazreku, A., O'Connor, P.J., Olsson, S.A., Ottesen, R.-T., Petersell, V., Plant, J.A., Reeder, S., Salpeteur, I., Sandström, H., Siewers, U., Steenfelt, A., Tarvainen, T., 2005. FOREGS Geochemical Atlas of Europe, Part 1: Background Information, Methodology and Maps. Geological Survey of Finland, Espoo, pp. 526. <http://weppi.gtk.fi/publi/foregsatlas/>.
- Scheib, A.J., Birke, M., Dinelli, E., The GEMAS Project Team, 2014. Geochemical evidence of aeolian deposits in European soils. *Boreas* 43 (1), 175–192. <http://dx.doi.org/10.1111/bor.12029>.

- Sinclair, A.J., 1976. Application of probability graphs in mineral exploration. In: Special Volume 4. Association of Exploration Geochemists, Toronto, pp. 95.
- Sinclair, A.J., 1983. Univariate analysis. Chapter 3. In: Howarth, R.J. (Ed.), Statistics and Data Analysis in Geochemical Prospecting. In: G.J.S. (Series Editor), Handbook of Exploration Geochemistry, vol. 2. Elsevier, Amsterdam, pp. 59–81.
- Tennant, C.B., White, M.L., 1959. Study of the distribution of some geochemical data. *Econ. Geol.* 54, 1281–1290.
- Thiombane, M., Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Cavaliere, M., De Vivo, B., 2017. Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model. *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.gexplo.2017.06.010>.
- Vercoutere, K., Fortunati, U., Muntau, H., Griepink, B., Maier, E.A., 1995. The certified reference materials CRM 142 R light sandy soil, CRM 143 R sewage sludge amended soil and CRM145R sewage sludge for quality control in monitoring environmental and soil pollution. *Fresenius J. Anal. Chem.* 352, 197–202.
- Violante, P., Adamo, P., 2000. pH determination. In: Violante, P. (Ed.), Official Methods of Soil Chemical Analysis. Francoangeli, Rome, pp. 10–13.
- Zuo, R., Wang, J., 2016. Fractal/multifractal modelling of geochemical data: a review. *J. Geochem. Explor.* 164, 33–41.
- Zuo, R., Wang, J., Chen, G., Yang, M., 2015. Identification of weak anomalies: a multi-fractal perspective. *J. Geochem. Explor.* 148, 12–24.
- Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Zuo, R., De Vivo, B., 2017. Exploring unielement geochemical data under a compositional perspective. *Appl. Geochem.* <http://dx.doi.org/10.1016/j.apgeochem.2017.10.003>.

Paper 4

Assessment of the behaviour of potentially toxic elements (PTEs) in soil from the Sarno River Basin through a compositional data analysis

Matar Thiombane, Josep-Antoni Martín-Fernández, Stefano Albanese, , Annamaria Lima, Angela Doherty, Benedetto De Vivo

Journal of Geochemical Exploration, Volume 195, December 2018, Pages 110-120



Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/jgexplo

Exploratory analysis of multi-element geochemical patterns in soil from the Sarno River Basin (Campania region, southern Italy) through compositional data analysis (CODA)

Matar Thiombane^{a,*}, Josep-Antoni Martín-Fernández^b, Stefano Albanese^a, Annamaria Lima^a, Angela Doherty^c, Benedetto De Vivo^d

^a Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse, Università degli Studi di Napoli "Federico II", Complesso Universitario Monte S. Angelo, Via Cintia snc, 80135 Naples, Italy

^b Dept. d'Informàtica, Matemàtica Aplicada i Estadística, University of Girona, Edifici Politècnica 4, Campus Montilivi, 17003 Girona, Spain

^c Auckland Regional Council, Bledisloe House, 24 Wellesley St, Auckland 1010, New Zealand

^d Pegaso University, Piazza Trieste e Trento 48, 80132 Naples & Benecon Scarl, Dip. Ambiente e Territorio, Via S. Maria di Costantinopoli 104, 80138 Naples, Italy

ARTICLE INFO

Keywords:

ArcGIS
Biplot
Centered log transformation (clr)
Compositional data analysis
Factor analysis
PTEs
Sarno Basin

ABSTRACT

The Sarno River Basin (south-west Italy), nestled between the Somma–Vesuvius volcanic complex and the limestone formations of the Campania–Apennine Chain, is one of the most polluted river basins in Europe due to widespread industrialization and intensive agriculture. Water from the Sarno River, which is heavily contaminated by the discharge of human and industrial wastes, is partially used for irrigation on the agricultural fields surrounding it. We apply compositional data analysis to 319 soil samples, collected during two field campaigns along the river course and throughout the basin, to determine the concentration and possible origin (anthropogenic and/or geogenic) of the elemental anomalies, including potentially toxic elements (PTEs).

The concentrations of 53 elements were determined using ICP-MS and, subsequently, log-transformed. Using hierarchical clustering, clr-biplot and a principal factor analysis, the variability and the correlations between a subset of extracted variables (26 elements) were identified. Factor score interpolated maps were then generated using both lognormal data (NDR) and clr-transformed data to better visualize the distribution and potential sources of the patterns in the Sarno Basin.

The underlying geology substrata appear to be associated with raised levels of Na, K, P, Rb, Ba, V, Co, B, Zr, and Li, due to the presence of pyroclastic rocks from Mt. Somma–Vesuvius. Similarly, elevated Pb, Zn, Cd, and Hg concentrations are most likely related to the geological and anthropogenic sources, the underlying volcanic rocks, and contamination from fossil fuel combustion associated with nearby urban centers. Interpolated factor score maps and the clr-biplot show a clear correlation between Ni and Cr in samples taken along the Sarno River, and Ca and Mg near the Solofra district. After considering nearby anthropogenic sources, the Ni and Cr are PTEs most likely originating from the Solofra tannery industry, while Ca and Mg correlate to the underlying limestone-rich soils of the area.

This study shows the applicability of the log-ratio transformations to these studies, as they clearly show relationships and dependencies between elements which can be lost when univariate and classical multivariate analyses are employed on raw and lognormal data.

1. Introduction

Environmental geochemistry investigates the impact of natural geochemical processes and anthropogenic activities on natural systems (e.g., rivers, lakes, soils, or forests), and on human health. These activities may lead to issues of ongoing anthropogenic global environmental change such as acid rain, ocean acidification, toxic metal

pollution, or water contamination, and may result in contamination of an area by potentially toxic elements (PTEs). A useful tool in determining the level and distribution of PTE contamination, and thereby identify areas where there may be a risk to the local ecology and to humans, is an environmental risk assessment (Aral, 2010; Health Canada, 2010a, 2010b; DEFRA, 2011).

The Sarno Basin, considered to be one of the most polluted areas of

* Corresponding author.

E-mail address: thiombane.matar@unina.it (M. Thiombane).

<https://doi.org/10.1016/j.jgexplo.2018.03.010>

Received 28 June 2017; Received in revised form 30 November 2017; Accepted 17 March 2018
0375-6742/ © 2018 Published by Elsevier B.V.

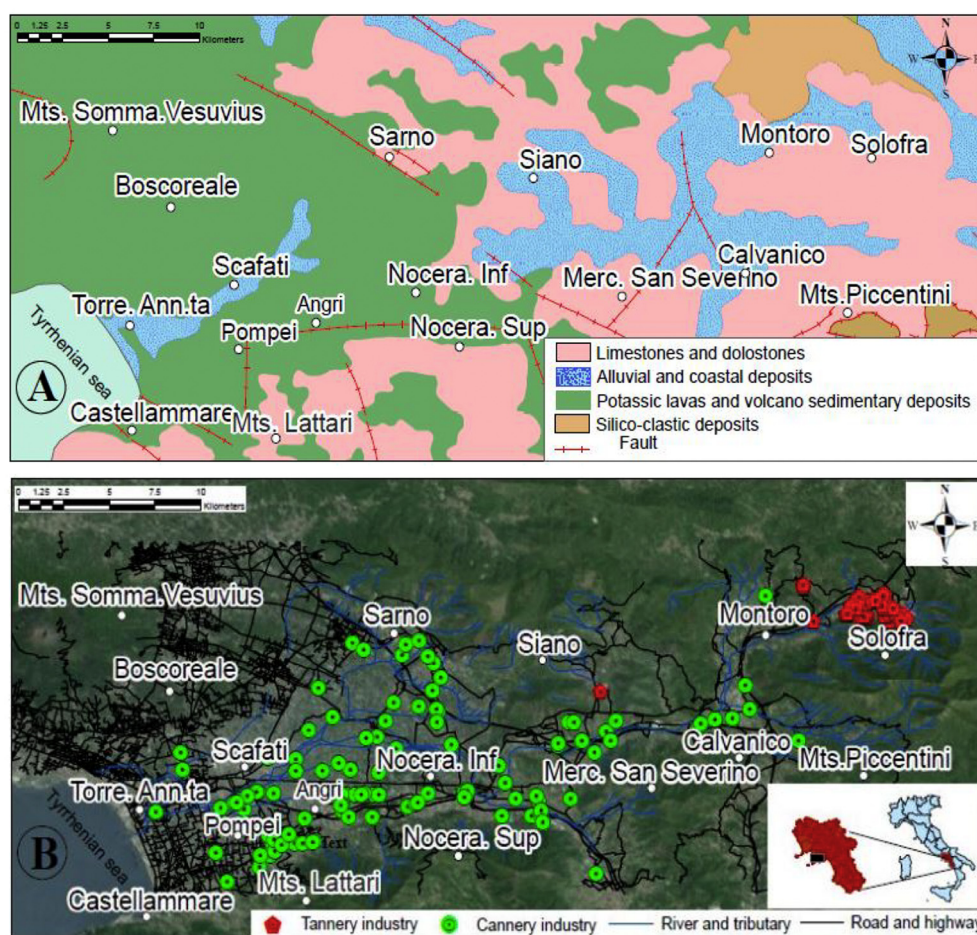


Fig. 1. The Sarno River Basin; (A) Simplified geological map of the Sarno Basin; (B) Map of land use and industrial activities.

Europe, is contaminated by PTEs from both anthropogenic sources including several different industries (Arienzo et al., 2001; Cicchella et al., 2014), as well as from the underlying geological substrata (Adamo et al., 2006). Statistical studies may be employed to define the characteristics of the PTEs in the Sarno Basin, including the definition of the background concentrations of these elements, the relationships between different chemical elements, and, in some cases, the determination of the extent of anthropogenic effects on the contamination status of the PTEs in the basin. It is understood that classical statistical methods based on Euclidean distance, and particularly those based on covariance structure, are not applicable to geochemical concentrations (Aitchison, 1986). This is mainly due to the positive character of geochemical data, and to the fact that they carry only information about the relative abundance of each individual component. To treat these compositional data using multivariate analysis, we must assume that the sample space is R^D , which can properly reconcile the correlation between elements. This was first recognized by Pearson (1897), who introduced the concept of spurious correlation for coefficients computed from raw compositions, which has since been studied by many authors, in particular by Chayes (1960), who tried to relate these issues to the singularity of the covariance matrix. Aitchison (1986) was the first to put forward a comprehensive alternative approach to classical statistics in order to avoid this problem, defining a new distance more suitable to compare compositions than the Euclidean distance. The most common approach to deal with concentration variables is now the use of log-ratio transformation (Aitchison, 1986; Aitchison et al., 2000; Aitchison and Egozcue, 2005) to express the compositions in terms of log-ratio coordinates. Three major transformations have been developed: additive log-ratio (alr), centered log-ratio (clr), both by Aitchison

(1986), and the isometric log-ratio (ilr) by Egozcue et al. (2003). This study uses the clr-transformation method, as the clr-coordinates are most appropriate to interpret the distances between elements (Palarea-Albaladejo et al., 2012; Thiombane et al., 2017), and to perform a dimensionality reduction through use of a biplot (Aitchison and Greenacre, 2002). Clr transformation was applied to the data in conjunction with clr-biplot compositional statistical transformation and Factor Analysis (FA), to distinguish the origin of the pollution in the Sarno River Basin. The main objectives of this multivariate investigation were:

- to determine the geochemical anomalies and sources of element patterns in the Sarno Basin.
- to distinguish the main vector and element associations into the environment.
- to use log transformation methods (lognormal and clr log-transformation) on the data to distinguish the potential anthropogenic or geogenic origin of the PTEs in the basin by applying clr-biplot and FA.
- to assess how the main lithological chemistry and human activities affect the Sarno Basin by using interpolated factors scores maps.

To facilitate better interpretation of the distribution patterns and the correlation between elements, the results of the clr transformed data were compared with the classical log transformed data by means of interpolated maps with ArcGIS software. Log-ratio transformations and clr-biplot were made using CoDaPack software (Comas-Cufí and Thió-Henestrosa, 2011), whereas the R Software was used to determine the correlation matrix and perform the FA.

2. Study area

Covering an area of 500 km², the Sarno River Basin is located in the Campania Region (southern Italy). The basin is bounded by the pyroclastic deposits of the Somma–Vesuvius volcanic complex to the north-west, the Tyrrhenian Sea to the west, the Picentini and Sarno Mountains to the east, and by Lattari Mountains and Sorrento Peninsula limestone reliefs to the south-west (Fig. 1A). Carbonate rocks belonging to the Campano–Lucanian and Abruzzese–Campanian carbonate platforms crop out in the Sarno River Basin including limestone outcrops around the Sarno and Lattari Mountains, Triassic dolomite, Lower Jurassic–Cretaceous limestone and dolomite limestone, and fractured and karstified Cretaceous limestone (De Pippo et al., 2006). Volcanic tuff, lapilli, and ash deposits from activity of the Somma–Vesuvius volcanic complex generally cover the calcareous-dolomitic rock of the Sarno Mountains (Fig. 1A). It is generally understood that the main soil characteristics of the study area are the result of geological phenomena (weathering, soil erosion) of the lithological substrate in and surrounding the Sarno Basin.

The Sarno River flows approximately 24 km through the basin, and is considered the most polluted river in Europe (Adamo et al., 2006; Albanese et al., 2013). Untreated domestic and agricultural effluent and solid and liquid industrial wastes are discharged directly into the river. It springs from the slopes of Mt. Sarno near the eponymous town, and empties into the Gulf of Naples (Tyrrhenian Sea) near Rovigliano Rock, between Castellammare di Stabia and Torre Annunziata (Arienzo et al., 2001). The river system traverses the Campanian provinces of Salerno, Avellino, and Naples, and collects water from the Solofrana and Cavaiola tributaries, as well as discharge from the municipalities of Montoro, Scafati and Mercato San Severino.

The population of the Sarno Basin is approximately 1,400,520 inhabitants, with an urban average density of about 1859 inhabitants per km² and a peak density along the coastal and some inland areas of over 2200 inhabitants per km² (ISTAT, 2013). The Sarno Basin is highly industrialized with > 160 tanneries located mostly in the Solofra area (Arienzo et al., 2001), canneries distributed along the length of the river, and the industrial area of Castellammare containing the Novartis pharmaceutical factories, and medium-scale industrial facilities producing paints, ceramics, and packaging food (Fig. 1B). The situation is aggravated by pervasive discharge of untreated agricultural and industrial waste directly into the river. The Sarno Valley is also home to very intensive agricultural activities, consisting mainly of tomato field horticulture in San Marzano, extensive orchards between Pompeii and Scafati, vineyards in the Boscoreale area, and chestnuts, greenhouse horticulture, floriculture, and crop production throughout the basin.

3. Materials and methods

3.1. Soil sampling and preparation

Soil sampling was carried out during two campaigns. During the first (known as the SAR–SAR campaign), topsoil samples were collected from a depth of 0–20 cm from the surface, at 36 sampling points of irrigated and flood-plain soil on average every 0.8 km along the main course of both the Sarno and Solofrana rivers in 2010. The second campaign (known as the SAR–SOB campaign) collected 283 soil samples from throughout the Sarno River Basin, with an average sampling density of approximately one sample per 2 km². In total, 319 samples were collected during the two campaigns (Fig. 2), following internationally recognized sampling methods (Plant et al., 1996; Salminen et al., 1998).

Each soil sample was collected as composites from five pits within an area of 100 m², and consisted of approximately 3 kg of soil, collected between 0 and 20 cm below the surface, stored in inert plastic bags. All samples were air-dried to prevent the volatilization of Hg, and sieved to collect 30 g of the < 2 mm fraction for chemical analysis.

3.2. Chemical analysis

Analyses were carried out at Acme Analytical Laboratories Ltd (now Bureau Veritas, Vancouver, Canada), an ISO 9002 accredited laboratory, within a 20 day time frame from when the samples were received, to the final delivery of analytical results. The samples were analyzed after an aqua regia extraction, by a combination of inductively coupled plasma atomic emission (ICP–AES) and inductively coupled plasma mass spectrometry (ICP/MS) for 53 elements (Ag, Al, As, Au, B, Ba, Be, Bi, Ca, Cd, Ce, Co, Cr, Cs, Cu, Fe, Ga, Ge, Hf, Hg, In, K, La, Li, Mg, Mn, Mo, Na, Nb, Ni, P, Pb, Pd, Pt, Rb, Re, S, Sb, Sc, Se, Sn, Sr, Ta, Te, Th, Ti, Tl, U, V, W, Y, Zn, and Zr) using Acme's Group 1F–MS package. A subsample of 15 g of the sieved < 2 mm soil fraction was digested in 90 ml aqua regia and leached for 1 h in a 95 °C water bath. After cooling, the solution was diluted to a final volume of 300 ml using a solution of 5% HCl. The sample weight to solution volume ratio was 1 g per 20 ml. The solutions were analyzed using a Spectro Ciros Vision emission spectrometer (ICP–AES) and a Perkin Elmer Elan 6000/9000 inductively coupled plasma emission mass spectrometer (ICP–MS). The accuracy and precision of the data was measured by comparison to known analytical standards. Calibration solutions were included at the beginning and end of each analytical run (a total of 40 solutions). Precision is ± 100% at the detection limit, and improves to better than ± 10% at concentrations 50 times the detection limit or greater (Table. 1).

3.3. Statistics and compositional data analysis (CoDA)

A composition is defined as a sample space of the regular unit D-simplex, S^D , that is a vector of D positive components summing up to a given constant k , set typically equal to 1 (proportions), 100 (percentages), or 10⁶ (ppm) by closure. It relates parts of some whole that carry relative information (ratios of components) whose sample space is the simplex.

$$S^D = \left\{ X = [x_1, x_2, \dots, x_D] \mid x_i > 0; \sum_{i=1}^D x_i = k \right\} \quad (1)$$

A statistical multivariate treatment should maintain the special nature of the compositional data and keep their correlations intact. Pearson (1897) showed that using Euclidean distance methods removes this context, leading to possibly misleading results. Aitchison (1986) was the first to put forward a comprehensive alternative approach to classical statistics in order to avoid this problem. By analogy to the log-normal approach, Aitchison (1986) projected the sample space of compositional data, the D -part simplex S^D , to real space, R^{D-1} or R^D , using log-ratio transformations. The additive log-ratio (alr) and the centered log-ratio (clr) transformation were introduced by Aitchison (1986) and the isometric log-ratio (ilr) transformation was introduced subsequently by Egozcue et al. (2003). Descriptions of the properties and processes of these transformations can be found in Pawłowsky-Glahn et al. (2015b).

3.3.1. The centered log-transformed data (clr)

The clr transformation for a composition $x = [x_1, x_2, \dots, x_D] \in S^D$ is the transformed data $y \in R^D$ with:

$$y = \text{clr}(x) = \left(\ln \frac{x_1}{g(x)}, \ln \frac{x_2}{g(x)}, \dots, \ln \frac{x_D}{g(x)} \right) \quad (2)$$

That is, each component is represented as a log-ratio of a central value by using the geometric mean of the composition x , $g(x) = [x_1 x_2 \dots x_D]^{1/D}$. The clr-transformed coordinates are always dependent on the geometric mean, so cannot be simply interpreted independently of other elements. Thus, just as in individual components, any single component of the clr-transformation will be associated with the “total”, or, the set of all considered components (Pawłowsky-Glahn et al., 2015a). The main advantages of the clr coefficients are: 1) the clr

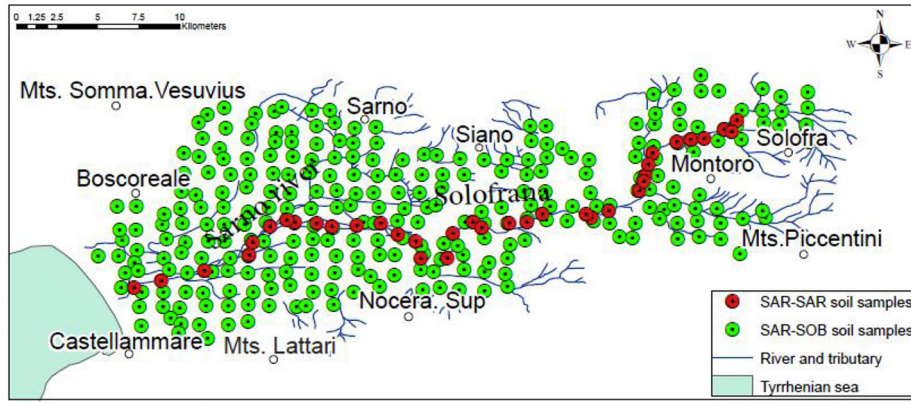


Fig. 2. Map of the study area with the location of soils sampling sites: SAR–SAR represents samples sites in along the length of the Sarno River; SAR–SOB samples sites are distributed across the Sarno Basin.

Table 1

Detection limits, accuracy and precision of the applied analytical method.

Element	Unit	Detection limit	Accuracy Δ (%)	Precision (%)
Al	%	0.01	0.4	1.2
Ca	%	0.01	2.5	1.2
Fe	%	0.01	0.7	1.8
K	%	0.01	2.5	3.8
Mg	%	0.01	1.5	1.4
Na	%	0.001	3.3	2.7
P	%	0.001	1.2	3.8
S	%	0.02	19.4	11.5
Ti	%	0.001	2.5	5.7
Ag	mg kg ⁻¹	0.002	6.5	10.9
As	mg kg ⁻¹	0.1	1.3	2.2
Au	mg kg ⁻¹	0.0002	8.2	21.9
B	mg kg ⁻¹	1	0.4	8.2
Ba	mg kg ⁻¹	0.5	0.3	1.7
Be	mg kg ⁻¹	0.1	0.7	2.3
Bi	mg kg ⁻¹	0.02	1.2	3.5
Cd	mg kg ⁻¹	0.01	1.4	5.1
Co	mg kg ⁻¹	0.1	2.3	3.2
Cr	mg kg ⁻¹	0.5	2.5	1.2
Cu	mg kg ⁻¹	0.01	1.6	3.7
Ga	mg kg ⁻¹	0.1	3.3	4.2
Hg	mg kg ⁻¹	0.005	2.4	6.3
La	mg kg ⁻¹	0.5	3.5	2.9
Mn	mg kg ⁻¹	1	1.5	2.9
Mo	mg kg ⁻¹	0.01	1.4	1.1
Ni	mg kg ⁻¹	0.1	0.8	2.7
Pb	mg kg ⁻¹	0.01	1.4	1.8
Sb	mg kg ⁻¹	0.02	1.3	2.3
Sc	mg kg ⁻¹	0.1	0.8	4.4
Se	mg kg ⁻¹	0.1	0.9	22
Sn	mg kg ⁻¹	0.1	0.5	3.4
Sr	mg kg ⁻¹	0.5	4.3	2.5
Te	mg kg ⁻¹	0.02	3.9	6.4
Th	mg kg ⁻¹	0.1	4.1	3.1
Tl	mg kg ⁻¹	0.02	2.2	4.1
U	mg kg ⁻¹	0.1	1.6	3.7
V	mg kg ⁻¹	2	2.3	3.2
W	mg kg ⁻¹	0.2	3.7	5.1
Zn	mg kg ⁻¹	0.1	1.8	2.2

coefficients translate perturbation and powering of compositions into ordinary sum and multiplication by a scalar of vectors of clr coefficients; and 2) classical Euclidean distance between vectors of clr coefficients is equal to the Aitchison distance of their corresponding compositions (Egozcue and Pawłowsky-Glahn, 2006). These two properties make the clr coefficients extremely useful. The Aitchison distance and the other metric properties including the scale invariance, permutation invariance, and sub-compositional coherence may be effectively related by the use of clr coordinates. Further discussion can be found in Barceló-Vidal and Martín-Fernández (2016).

The main disadvantage of the clr transformation is that the clr covariance matrix is singular. Despite this complication, the relationship between the chemical elements can be freely analyzed using the clr-biplot (Aitchison and Greenacre, 2002), and it is for this reason that clr coordinates were used in our analysis and statistical computations.

One of the best methods to display the variable associations and their main behaviour in multivariate analysis is by using a dendrogram arising from hierarchical clustering. Therefore a hierarchical cluster was built based on lognormal data and clr logratio transformation (Fig. 3) to better visualize several clear variable associations that can be explained from knowledge of their geochemical behavior, or nearby anthropogenic processes in the survey area.

3.3.2. Clr-biplot

The most common graph to portray the relationship between variables is the biplot (Gabriel, 1971; Kempton, 1984). The biplot is a 2D graphical representation based on the singular value decomposition (Eckart and Young, 1936), where the individuals or sample observations are expressed as dots and the variables as rays. It permits to relate the correlation and the association between variables, and the

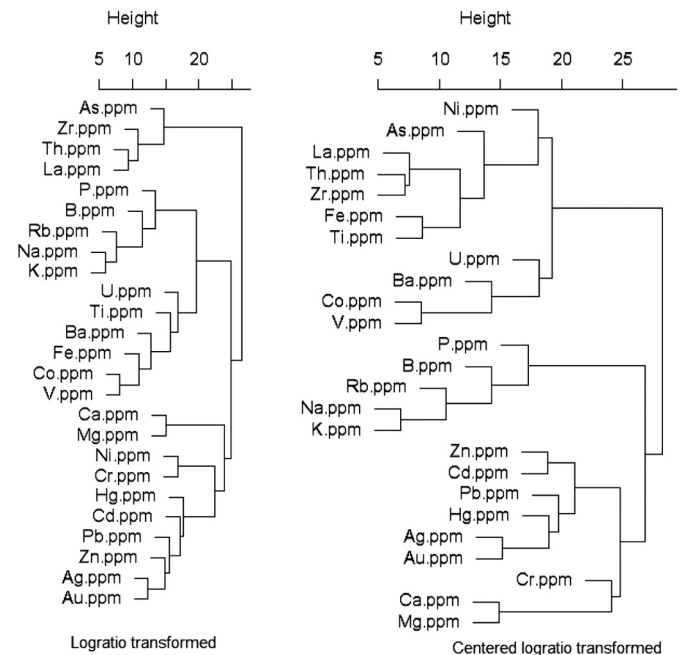


Fig. 3. Dendrograms for lognormal data (left plot) and clr-transformed data (right plot) of the selected 26 variables.

Table 2

Statistical parameters and descriptive statistics of the 26 elements included in the final variables matrix after varimax orthogonal rotation from soil sediment samples (n = 319) of the Sarno River Basin (values presented in mg/kg).

Variable	Unit	Samples	Minimum	Maximum	Mean	Median	Std deviation	Skewness	Kurtosis
Ag	mg/kg	319	0.03	0.63	0.14	0.12	0.09	2.55	9.03
As	mg/kg	319	5.82	111.50	14.65	13.10	8.21	8.68	93.18
Au	mg/kg	319	0.0003	0.21	0.01	0.01	0.02	7.98	91.35
B	mg/kg	319	05.00	46.00	21.86	22.00	7.43	0.19	−0.34
Ba	mg/kg	319	294.50	1034.90	645.50	643.00	155.25	0.08	−0.51
Ca	mg/kg	319	4300.00	127,700.00	38,739.00	31,500.00	23,874.00	1.39	1.85
Cd	mg/kg	319	0.15	11.06	0.62	0.54	0.82	11.26	134.48
Co	mg/kg	319	06.00	16.60	12.14	12.30	2.19	−0.31	−0.33
Cr	mg/kg	319	5.30	808.40	35.96	16.80	76.65	5.99	43.19
Hg	mg/kg	319	0.01	0.61	0.09	0.06	0.07	2.77	12.36
K	mg/kg	319	4000.00	47,600.00	23,907.00	23,300.00	10,705.00	0.02	−1.16
Fe	mg/kg	319	15,700.00	41,200.00	29,865.52	30,100.00	4222.02	−0.45	0.25
La	mg/kg	319	19.70	86.70	46.00	44.70	11.37	0.63	0.55
Li	mg/kg	319	8.40	45.50	21.09	19.80	7.10	0.85	0.42
Mg	mg/kg	319	3300.00	46,900.00	10,317.00	8700	5557.00	2.83	10.48
Na	mg/kg	319	460.00	15,360.00	6444.00	6640.00	3583.00	−0.04	−1.28
Ni	mg/kg	319	8.50	29.40	15.94	15.60	1.32	1.32	3.40
P	mg/kg	319	460.00	6170.00	2720.00	2680	1129.00	0.32	−0.27
Pb	mg/kg	319	25.41	585.16	71.72	63.47	41.13	6.90	76.36
Rb	mg/kg	319	51.20	259.40	171.31	174.20	49.42	−0.27	−0.91
Th	mg/kg	319	3.60	47.80	14.14	13.20	5.27	1.91	7.62
Ti	mg/kg	319	620	2670.00	1632.00	1620.00	342.00	0.17	0.26
U	mg/kg	319	1.80	23.70	4.98	4.60	2.16	3.16	19.90
V	mg/kg	319	43.00	131.00	91.06	90.00	17.12	−0.06	−0.32
Zn	mg/kg	319	60.50	1115.50	160.47	140.60	93.58	4.22	34.42
Zr	mg/kg	319	7.40	269.80	37.59	28.80	29.12	4.03	23.64

construction of a smooth distribution of elements depending of their characteristics. When the biplot is used with clr log-transformed data, it facilitates the analysis of all geochemical variables (Aitchison and Greenacre, 2002; Otero et al., 2005). In addition, the lengths of the rays and the angles between rays and between links may be interpreted in statistical terms (Pawlowsky-Glahn et al., 2015b). The coordinates of the samples represented in the clr-biplot are the coordinates of the centered data. Consequently, the Euclidean distance between two samples in the clr-biplot is an approximation of the Aitchison distance between the corresponding compositions in the simplex. The squared length of a ray is proportional to the clr-variance of the corresponding chemical element. The squared length of the link between two vertices of rays is approximately equal to the variance of the log-ratio of the parts corresponding to the clr-rays. In this sense, the closer the vertices are, the more proportional the concentrations of the chemical elements are (Aitchison and Greenacre, 2002). The cosine of the angle between two links approaches the correlation coefficient between the corresponding simple log-ratios. Consequently, the orthogonality of the links suggests that there is no correlation of the log-ratios; it is not possible to estimate the covariance matrix in order to obtain robust counterparts of loadings and scores for the construction of a robust compositional biplot (Filzmoser and Hron, 2011).

3.3.3. Interpolated factor score maps

The interpolated factor association maps have been obtained using ArcGIS 10.2 Spatial Analyst software, using Inverse distance weighted (IDW) as the interpolation method. This method assumes that that values of neighboring observations contribute more to interpolated values of distant observation even if it ignores the variance of values (Cheng, 2008). A distance of 1.5 km was chosen as the linear-weighted combination distance between sample points. Moreover, the structural geology, the environment, and landscape morphologies and characteristics have been taken into account to better visualize spatial patterns of elements associations into the survey area. The different factors score maps obtained were studied and interpreted in accordance with their hypothetical origin (geogenic, anthropogenic or mixed).

3.4. Factor analysis (FA)

Factor analysis is a statistical method that characterizes different groups of chemical elements with approximately similar geochemical patterns (Miesch Programs, 1990). Multivariate statistical analysis was applied to the lognormal of raw data (NRD) and the clr log-transformed data in order to make in evidence the suitable interrelationships between variables clr log-transformed. One objective of this study is to determine the relationships between the 26 variables in the 319 samples of this study using factor analysis. To determine the distribution and the correlation structure of the variables, we employed multivariate statistical analysis by using the principal factor analysis, which clearly identifies unique factors that have a completely different behavior to the majority of all other factors. This method aids another objective of this study, the determination of the geogenic or/and anthropogenic source of the variables (Reimann and De Caritat, 2005). To better visualize the correlation between the variables, the number of elements considered was reduced based on 3 criteria: 1) remove elements which concentrations are lower than the limit of detection (LOD); 2) choose arbitrary mostly two representative elements (e.g., Rare Earth Elements; REE), which are geochemically congruent; 3) choose elements with a communality of extraction higher than 0.5 (50%) and/or a common variances < 0.5. The factor association maps were carried out using lognormal data and the clr log-transformation.

A varimax orthogonal rotation was employed as it not only represents how the variables are weighted for each factor, but also the correlation between the variables themselves and the factor. As it is an orthogonal rotation that minimizes the number of variables that have high loadings on each factor, it simplifies the transformed data matrix, which aids in data interpretation. Details of the 26 elements included in the final variables matrix (Ag, As, Au, B, Ba, Ca, Cd, Co, Cr, Fe, Hg, K, La, Li, Mg, Na, Ni, P, Pb, Rb, Th, Ti, U, V, Zn, Zr) are shown in Table 2.

Based on the multivariate analysis, four factor associations (F1, F2, F3, and F4) were derived for both the lognormal data and the clr-transformed data (Table 3). These factor associations were considered and interpreted in accordance with the presumed origin, or reason for the variable association, (i.e., geogenic or anthropogenic sources).

Table 3

Varimax-rotated factor (four-factor model) of 26 variables for 319 soil samples from the Sarno basin; table shows the rotated component matrix of the lognormal data (left table) and the clr transformed data (right table). Bold entries: loading values over $|0.50|$.

Variables	Factors				Communalities		Clr (variables)	Factors				Communalities	
	F1	F2	F3	F4	Initial	Extraction		F1	F2	F3	F4	Initial	Extraction
Ag	0.10	−0.18	0.73	0.32	1.00	0.68	Ag	0.22	−0.71	−0.08	0.22	1.00	0.61
As	−0.22	0.36	0.12	0.64	1.00	0.60	As	−0.74	0.23	0.26	0.10	1.00	0.67
Au	0.04	−0.07	0.62	0.05	1.00	0.59	Au	0.28	−0.75	−0.03	0.03	1.00	0.65
B	0.77	−0.22	0.22	−0.14	1.00	0.70	B	0.82	−0.06	−0.08	−0.17	1.00	0.72
Ba	0.89	0.23	0.09	−0.05	1.00	0.86	Ba	0.72	0.36	0.36	0.07	1.00	0.79
Ca	−0.22	−0.68	0.24	−0.20	1.00	0.61	Ca	0.25	−0.18	−0.86	0.01	1.00	0.83
Cd	−0.03	0.03	0.56	−0.03	1.00	0.52	Cd	−0.07	−0.64	−0.04	0.17	1.00	0.54
Co	0.79	0.40	−0.11	0.27	1.00	0.87	Co	0.39	0.67	0.30	0.46	1.00	0.90
Cr	−0.16	−0.10	0.10	0.83	1.00	0.73	Cr	−0.18	−0.27	−0.16	0.72	1.00	0.65
Fe	0.44	0.77	−0.17	0.26	1.00	0.88	Fe	−0.14	0.81	0.27	0.37	1.00	0.88
Hg	−0.06	−0.19	0.59	0.19	1.00	0.51	Hg	−0.15	−0.71	−0.03	−0.08	1.00	0.54
K	0.95	−0.11	0.00	−0.14	1.00	0.94	K	0.95	0.11	0.11	−0.07	1.00	0.94
La	−0.33	0.85	−0.08	0.15	1.00	0.87	La	−0.74	0.52	0.22	0.04	1.00	0.87
Li	−0.58	0.68	−0.13	0.23	1.00	0.87	Li	−0.84	0.39	0.01	0.13	1.00	0.86
Mg	−0.43	−0.52	−0.01	−0.06	1.00	0.56	Mg	−0.10	0.07	−0.87	0.14	1.00	0.79
Na	0.87	−0.17	0.02	−0.20	1.00	0.83	Na	0.92	0.08	0.04	−0.12	1.00	0.87
Ni	0.05	0.13	0.12	0.79	1.00	0.65	Ni	−0.20	0.19	−0.20	0.84	1.00	0.83
P	0.73	−0.19	0.39	0.01	1.00	0.72	P	0.79	−0.38	0.06	0.05	1.00	0.77
Pb	−0.01	−0.04	0.70	−0.06	1.00	0.50	Pb	0.02	−0.69	0.06	0.02	1.00	0.59
Rb	0.93	0.01	−0.03	−0.11	1.00	0.87	Rb	0.84	0.32	0.17	−0.07	1.00	0.84
Th	−0.28	0.87	−0.15	0.00	1.00	0.86	Th	−0.58	0.63	0.29	−0.01	1.00	0.82
Ti	0.19	0.79	−0.26	0.03	1.00	0.72	Ti	−0.24	0.79	0.22	0.13	1.00	0.75
U	0.32	0.65	0.16	−0.17	1.00	0.58	U	0.09	0.24	0.72	−0.16	1.00	0.60
V	0.81	0.46	−0.16	0.09	1.00	0.89	V	0.41	0.74	0.32	0.24	1.00	0.88
Zn	0.31	−0.14	0.64	0.04	1.00	0.52	Zn	0.39	−0.68	−0.07	0.13	1.00	0.63
Zr	−0.51	0.73	−0.14	−0.12	1.00	0.83	Zr	−0.76	0.56	0.12	−0.06	1.00	0.90
Eigenvalues	7.41	6.15	2.96	1.55	–	–	Eigenvalues	9.30	6.34	2.24	1.62	–	–
% variance	28.51	23.67	11.39	5.98	–	–	% variance	35.76	24.40	8.62	6.22	–	–
Cumulative %	28.51	52.18	63.57	69.54	–	–	Cumulative %	35.76	60.16	68.77	74.99	–	–

4. Results and discussion

4.1. Hierarchical clustering of lognormal versus clr-transformed data

Fig. 3 presents a hierarchical clustering of 26 variables lognormal data versus clr log-transformed data, based on the average linkage between elements method. No clear interrelationships and associations between groups of variables are visible in the hierarchical clustering on log-transformed data (Fig. 3, left plot). The Ca–Mg and Ni–Cr groups are correlated into the same cluster, and in fact, it seems that possible associations of variables can be simplified or/and reduced using the lognormal data. Performing hierarchical clustering on clr log-transformed data (Fig. 3, right plot) highlights the correlation between variables into a Euclidean geometry which clearly identifies the association of groups of elements in different variables. The variables cluster smoothly based on their most likely geochemical origins or interactions, such as the association of Ca–Mg arising from rudist limestone and carbonate sediments between Solofra and Montoro, and the association between Na–K–P–B and U–Ba–Co–V resulting from pyroclastic rocks of the Mt. Somma-Vesuvius. In addition to the background geology, human activities affecting the survey area can be identified such as through the element association of Ni–Cr and Pb–Zn–Cd–Hg relating to industrial activities such as the famous tanneries of the Solofrana district (source of Cr and Ni), metal processing operations, and the combustion of petroleum products. It is clear that the cluster analysis was able to distinguish particular groups and associations of variables, allowing us to draw conclusions as to their main geogenic and anthropogenic origins in survey area.

4.2. Generation of clr-biplot

The clr biplot (Fig. 4) explains almost all Euclidean variability (~70%) using only 26 of the total 52 elements analyzed in the first

three Principal Components (PC1, PC2, and PC3, accounting for 35.5%, 22.5%, and 10.2%, respectively). Applying this geostatistical computation to the Sarno database displays associations between elements that may reflect their sources, whether related to the underlying geology, or human activity, or a mixture of the two. Following projection into x, y coordinate space (Fig. 4, left plot) we may distinguish five main associations of variables, based on the length of the vertex, their vicinity to each to another and the position between dots (observations or samples) and rays.

The Ca–Mg association (A1) shows both variables with high communalities (the length of the vector) and high correlation coefficient ($r = 0.76$). These two elements are commonly found in the carbonate rocks of the study area, such as the silico-clastic deposits in the rudist limestone and carbonate sediment between Solofra and Montoro (see Table 3, F3).

Na–K–P–B–Rb–Ba association (A2) is dominated by a high communality of the Na variable and the vicinity of their rays. Within this group, the correlation coefficient between Na and K is higher ($r = 0.91$), followed by Na–P ($r = 0.69$), B–P ($r = 0.67$) and P–K ($r = 0.62$). These associations probably reflect the potassic and ultrapotassic rocks formations that occur throughout the majority of the basin, related to the lava and pyroclastic volcanic activity of Mt. Somma-Vesuvius.

Mercury (Hg) presents a high length of communality contributing to the (A3) association with Zn, Pb, Cd, Au, and Ag (see F2 of FA table). This association may be divided in two sub-groups, named Ag–Au ($r = 0.65$) and Zn–Cd–Pb–Hg sub-associations. This association contains both short vectors (Zn, Cd, and Pb are poorly represented) and long vectors (Au, Ag, and Hg). This may, in fact, be representative of the effects of both the geogenic source (volcanic material) and anthropogenic source (the high level of pollution in the basin related automobile gasoline combustion and industrial activities) (Albanese et al., 2013).

As–U–Th–Zr–La–Li (A4) and Fe–Ti–Co–V (A5) association vectors

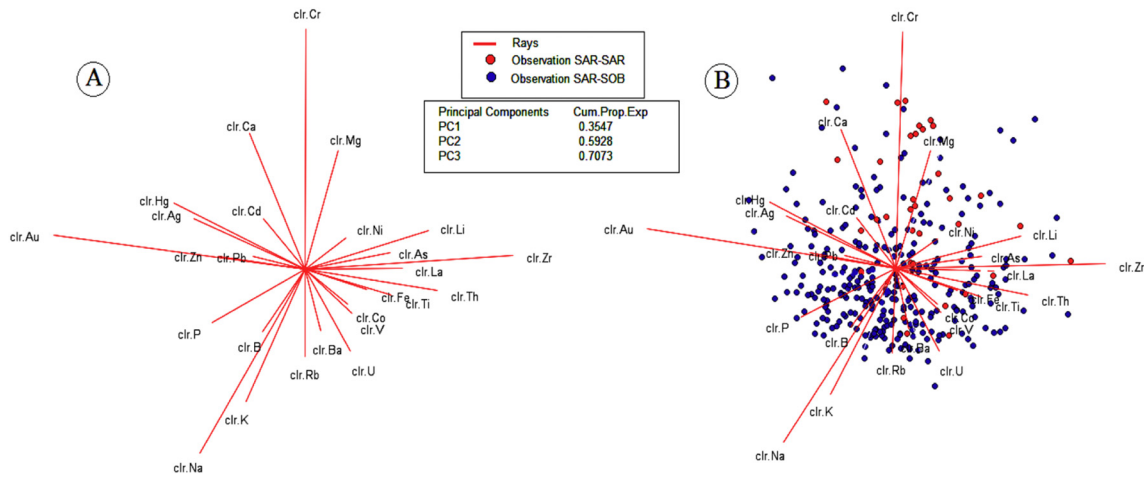


Fig. 4. Clr-biplot of the Sarno Basin soil samples. The left plot (A) is displayed without the individual sample observations, while the right plot (B) shows association between variables and the observations.

display generally the same length of communality, with the exception of Zr (which is significantly greater: 78.8%). Fe is positively correlated with Ti, V, and Co, with correlation coefficients of $r = 0.79$, $r = 0.78$, and $r = 0.65$, respectively. In the A4 association, a similar trend is seen with Th correlating strongly with Zr, La, Li, and U, with correlation coefficients of $r = 0.87$, $r = 0.85$, $r = 0.79$, and $r = 0.63$, respectively. In considering the most likely potential origin of these associations, the A4 association reflects the pyroclastic material covering hilly and mountain areas surrounding the Sarno River Basin, while A5 is most likely related to the predominantly magmatic rocks of Mt. Somma-Vesuvius, with their primary compositions of coexisting Fe–Ti oxides (Cicchella et al., 2014; De Pippo et al., 2006).

Ni and Cr (A6) are modestly correlated ($r = 0.58$), and Cr presents a communality value (84.12%). They might reflect a mixing of the main anthropogenic industrial activities of the river basin (i.e., the tannery district of Solofra, metal processing operations, and petroleum combustion). The blue dots (Fig. 4, right plot) represent the SAR–SOB sampling campaign across the entire basin, while the green dots represent samples collected along the river during the SAR–SAR campaign. It is of note that the SAR–SAR samples are generally correlated to the A6 (Ni–Cr) and A1 (Ca–Mg) vertices. This may suggest that the geochemistry associated with soils in contact with the river, and therefore the river waters themselves, are affected by wastewater from the local industrial activity such as the use of Cr to treat animal skins in the aforementioned tannery industry of the Solofra district.

To clarify these hypotheses, another clr-biplot (Fig. 5) was created

based on the amalgamation of the 26 variables in six possible associations (clr-amalgA1, clr-amalgA2, clr-amalgA3, clr-amalgA4, clr-amalgA5, and clr-amalgA6).

The amalgamation operation is frequently used to reduce the number of individual parts in compositional dataset, but it is a non-linear operation in the simplex with the geometry based on the Aitchison geometry (Mateu-Figueras and Daunis i Estadella, 2008). If the D parts of a composition are separated into $C \leq D$ mutually exclusive and exhaustive subsets, and the components of each subset are added together, the resulting C -part composition is termed an amalgamation (Aitchison, 1986). For example, from the 6-part composition ($x_1; x_2; x_3; x_4; x_5; x_6$) we may obtain the following 3-part amalgamation ($Am_1 = x_1 + x_2$; $Am_2 = x_3 + x_4$; $Am_3 = x_5 + x_6$). Elements are assembled into groups which may better explain the underlying associations. However, as pointed out in Pawlowsky-Glahn et al. (2015b, page 52) “amalgamation of parts is problematic when the amalgamated composition is compared with the original one”. In this work, we use the amalgamation algorithm to clarify the possible behavior of the groups of elements and their main associations. Based on this assertion, the clr-biplot (Fig. 5) was created comprising the amalgamation of the five proposed associations (A1, A2, A3, A4, and A5) based on the compositional clr-biplot (Fig. 4).

The amalgamation of the 26 elements into five association groups reveals important information.

The clr-amalgA1 (Ca, Mg) and clr-amalgA2 (Na, P, K, B, Rb) may be independent of each other since their rays form 90° angle. This could be

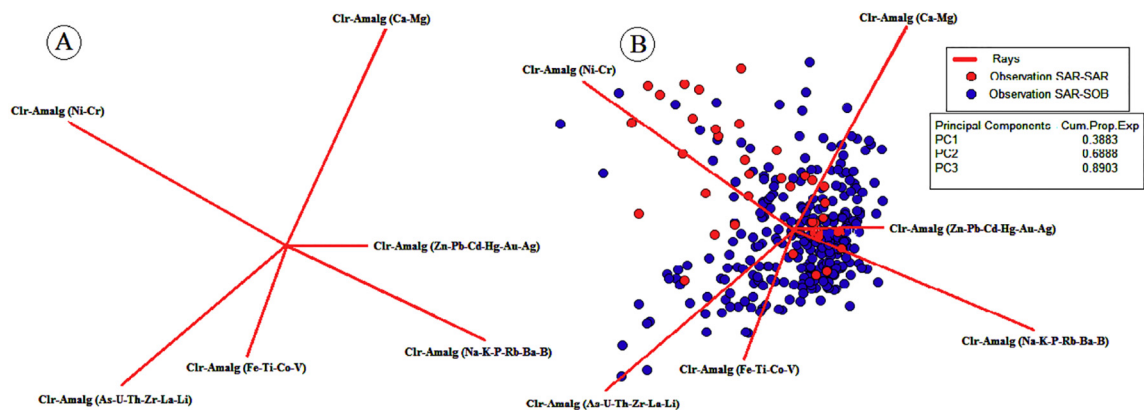


Fig. 5. Clr-biplot of the Sarno Basin soil samples; (A) The clr biplot of amalgamated variables without individual sample observations and (B) with the individual samples included. The amalgamation operation reduced the number of variables based on the initial compositional associations (Fig. 4). This clarifies the mixed anthropogenic and geological origin of the variables in the survey area.

due to the different geological substrates of the survey area, confirming the idea that the A1 and A2 associations represent material derived from different lithologies. The same scenario is displayed between clr-amalg (A6) and clr-amalg (A2) which their vertex form a 180° angle and in opposite directions. Clr-amalg (A4: As–U–Th–Zr–La–Li) and clr-amalg (A5: Fe–Ti–Co–V) share closely spaced rays but have differing communalities. This may indicate that they have the same geolithological sources, but result from different volcanic eruptions of Mt. Somma-Vesuvius. Based on the geological background occurring within the basin, A4 and A5 are most likely related to magmatic rocks (pyroclastic rocks) that occur the study area. Nevertheless, we suggest that the different communalities between the two associations may reveal the influence of sub-surface geological phenomenon (magmatism, crystallization) and/or external conditions (weathering, erosion) in the survey area. The high correlation of the Ti–Fe–Co–V association in the study area might be explained by the co-precipitation and absorption phenomena, due to the presence of Fe and Ti hydroxides in the soils developed from pyroclastic covers and volcano-sedimentary rocks.

The *clr-amalgA6* (Ni–Cr), and *amalgA1* (Ca–Mg) form an almost 90° angle; this may reflect a different origin, with one (A1) related to the geological sub-strata, and the other (A6) related to anthropogenic activities. Furthermore, A6 may be independent of the other groupings all together, reinforcing a non-geological origin, being related to activities of the tannery industries in the survey area (Albanese et al., 2013). The biplot (Fig. 5, right plot) confirms this interpretation, where the SAR–SAR observations, samples collected along the Sarno River are clearly associated with the *clr-amalgA6* ray (Ni–Cr vertex). Moreover, the correlation coefficient of the two variables Ni and Cr is $r = 0.58$ (positively correlated). This information indicates that the Ni–Cr contamination is associated with the river and/or is related to a matrix that discharged into the river. It can be confidently attributed to contamination from the tannery industries (Solofra and Montoro districts), which discharge their wastes directly into the Sarno river and are rich in Cr and enriched in Ni.

4.3. Production of interpolated factor association maps

To show the applicability of clr transformation applied to the factor analysis more clearly, two types of interpolated factor score maps were created, one using the lognormal data (NRD) and the other using the clr transformed data (CLR). This allows comparison between the two types of data, and will highlight the benefit of using clr-transformed data in showing the distribution of variables in the survey area, and their possible origin. As described above, the different factors obtained from factor analysis were considered and interpreted in accordance with their presumed origin (natural, anthropogenic, or mixed source). A total of 4 factors can be used to account for 69.54% and 74.99% of total cumulative variability in the NRD, and clr-transformed data, respectively (Table 3). This observation confirms that the clr-transformed data more clearly reflects the relationships, associations, and information contained in the original variables. Based on the results of factor analysis, interpolated maps were developed to further clarify the presumed origin of these elements, using frequency space-method (Figs. 6–9). The associations of the four-factor model using lognormal data are: F1: Na, K, P, Rb, Ba, V, Co, B, –(Zr, Li); F2: Th, La, Fe, Ti, Zr, Li, U, –(Ca, Mg); F3: Pb, Zn, Cd, Hg, Au, Ag; F4: As, Ni, Cr. For comparison, the associations of the others four-factor model using clr-transformed data are: F1: Na, K, P, Rb, Ba, B, –(Zr, Li, As, Th, La); F2: Th, La, Ti–Zr, Fe, Co, V, –(Pb, Zn, Cd, Hg, Au, Ag); F3: U, –(Ca, Mg); and F4: Ni, Cr.

Elements with loadings over $|0.5|$ are considered representative members of each association factor that results from the chosen factor model, permitting the construction of these interpolated factor association maps.

Fig. 6 and Table 3 display factor scores of variables in the F1 factors of the NRD and clr-transformed data range from -2.79 to 1.80 , and -3.5 to 2.1 , with a percentage of variability of 28.51% and 35.76%,

respectively. The factor map (Fig. 6) shows (particularly for the elements Na, K, and P) elevated factor scores near the slopes of Mt. Somma-Vesuvius, which decreases gradually towards the inner basin. This may be clearly attributed to the occurrence of igneous units and volcanic soils. These units are dominated by potassic and ultrapotassic lavas and pyroclastic material, which occupies the north-west Sarno Basin. We can therefore assume, that these maps reflect the concentrations of elements in lithological background of the Sarno Basin due to the pyroclastics rocks of the volcanic soils.

Fig. 7 and Table 3 display factor scores of variables in the F2 factor of the NRD and clr-transformed data range from -2 to 5.9 and -3.57 to 2.62 , with a percentage variability of 23.66% and 24.40%, respectively. The construction of these two interpolated factor maps clarifies the behavior of these elements and their correlations in the basin. In fact, Fig. 7 (upper map) identifies the areas containing the highest factor scores (ranging from 3.95 to 6.0 NDR F2) in the Solofra and Montoro areas and the lowest factor scores are mapped at the slope of the Mts. Picentini and Mts. Lattari. This supports the notion of a lithological background origin to these elements in this area, which is dominated by silicoclastic deposits of the Solofra and Montoro areas, and limestone and dolostone of the Picentini-Taburno unit and Lattari Mountains for Ca and Mg association. The factor score map (Fig. 7, lower) shows the lowest factor scores ranging from -3.57 to -1 and the Pb, Zn, Cd, Hg, Au, and Ag elemental association correspond to the locations of the most urbanized areas (Castellammare di Stabia, Scafati, Montoro, Calvanico, Nocera, and Solofra) in the Sarno basin. The anthropogenic sources of these variables include the high rates of fossil fuel combustion and vehicle emissions, releasing mostly Pb and Zn into the air and soil (De Pippo et al., 2006; Rossini and Fernández, 2007).

Fig. 8 and Table 3 display factor scores of variables in the F3 factor of the NRD and clr-transformed data range from -1.52 to 5.73 and -3.19 to 3.58 (Fig. 8), with a percentage variability of 11.39% and 8.62%, respectively. The location of hotspots seen in the upper map of Fig. 8 confirms the likely anthropogenic sources of the Pb, Zn, Cd, Hg, Au, Ag elemental associations, which correspond to the location of the main urban centres, while the lower map of Fig. 8 displays the highest factor scores (ranging from 2.5 to 3.58) coinciding with the silicoclastic deposits surrounding the Solofra and Montoro district. Fine-sized solid particles with a charged substrate, such as clay minerals, play an important role in the sorption of trace elements. In fact, U is a highly soluble element and can be easily dissolved, transported and precipitated within sedimentary deposits by subtle changes in oxidation conditions reflecting adsorption and co-precipitation phenomena.

Fig. 9 and Table 3 display factor scores of variables in the F4 factor of the NRD and clr-transformed data range from 2.48 to 9.85 and 2.09 to 4.05, with a percentage variability of 5.19% and 6.98%, respectively. The highest factor scores values (> 5) for NRD F4 (Fig. 9, upper) are found near the Solofra and Montoro districts, while high factor scores (> 2.01) for CLR F4 (Fig. 9, lower) are located associated with the townships of Solofra, Montoro, Mercato San Severino, Nocera, Scafati, and Castellammare di Stabia (Fig. 9) districts. In fact, all these areas are crossed by the Sarno River or the Solofrana tributaries, which are most likely control the movement of these particles through the study area. The clr transformed data shows a strong correlation of the Cr and Ni elemental all along the Sarno River and its tributaries, which the underlying geology cannot account for. Therefore, it is instead more likely introduced into the environment through anthropogenic activities. Indeed, it is most likely introduced by the release of wastewater from the tannery district of Solofra. This is supported in the literature, where the tannery industrial wastewater is considered to be a major source of pollution, due to presence of chlorophenol and Cr combined with Ni (Tariq et al., 2005). Chromium salts (particularly Cr sulphate) are the most widely used compounds associated with the tanning of skin, while Ni is released from additional tanning agents used during the process and metals industrial activities and the combustion of petroleum products.

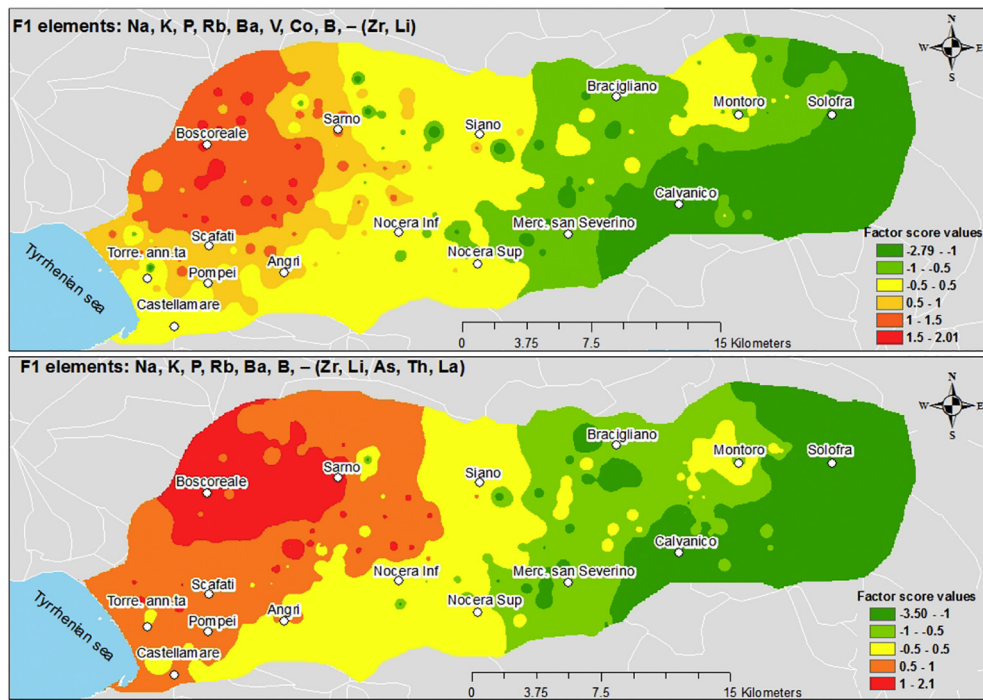


Fig. 6. Interpolated factor score maps showing the elemental association in soils of the survey area based on factor analysis results; (upper) factor scores map F1 (Na, K, P, Rb, Ba, V, Co, B, - (Zr, Li)) using lognormal data, and (lower) factor scores map F1 (Na, K, P, Rb, Ba, B, - (Zr, Li, As, Th, La)) using clr transformed data.

5. Finals remarks and conclusions

This study has been the first to use clr-transformed data based on Aitchison distance to interpret the distribution and correlation of multi-elements in the Sarno River Basin, instead of conventional multivariate analysis on raw data (e.g., Albanese et al., 2013; Cicchella et al., 2014). It shows the applicability of the compositional data analysis, which relates the unique relationships and dependencies between elements that can be lost when univariate and classical multivariate analyses are employed on lognormal transformed data. The combination of use clr-

transformation data, hierarchical clustering, principal factor analysis and the construction of interpolated factor score maps allow us to draw the following conclusions:

- It is possible to discern that the anomalies seen in the geochemical distribution of elements in the Sarno Basin have been affected to some degree by both the lithological chemistry (e.g., Pb, Zn, Cd, and Hg in pyroclastic rocks), and the industrial human activities (e.g., Ni and Cr in the tanning and metal industry discharges and emissions) using the clr-transformed coordinates, i

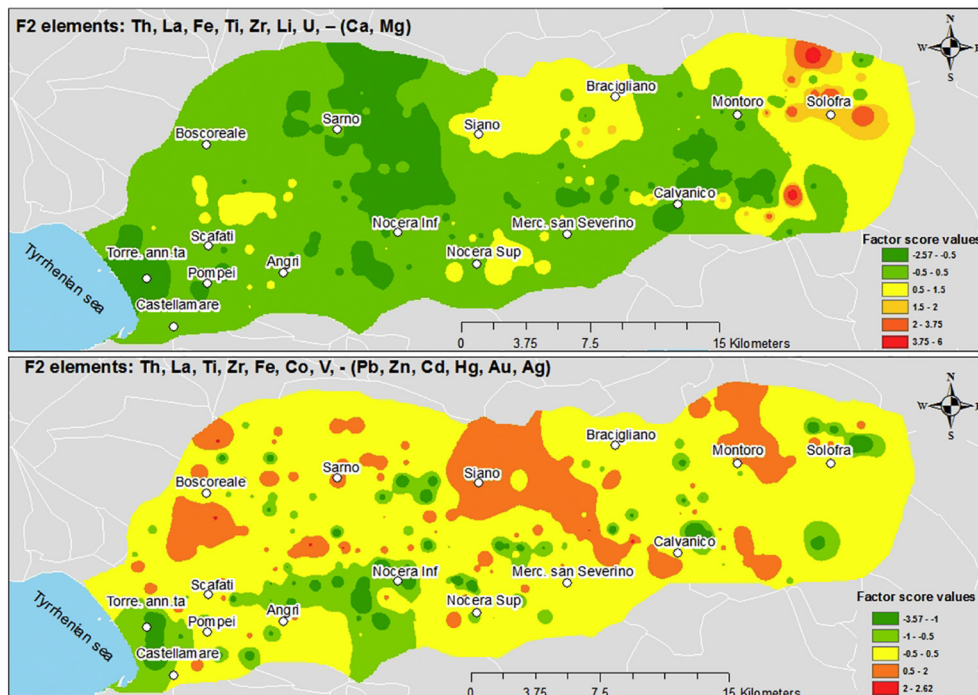


Fig. 7. Interpolated factor score map showing the elemental association in soils of the survey area based on factor analysis results. Upper figure shows the factor scores map F2 (Th, La, Fe, Ti, Zr, Li, U, - (Ca, Mg)) using the lognormal data, and the lower one displays factor scores map F2 (Th, La, Ti-Zr, Fe, Co, V, - (Pb, Zn, Cd, Hg, Au, Ag)) using clr transformed data.

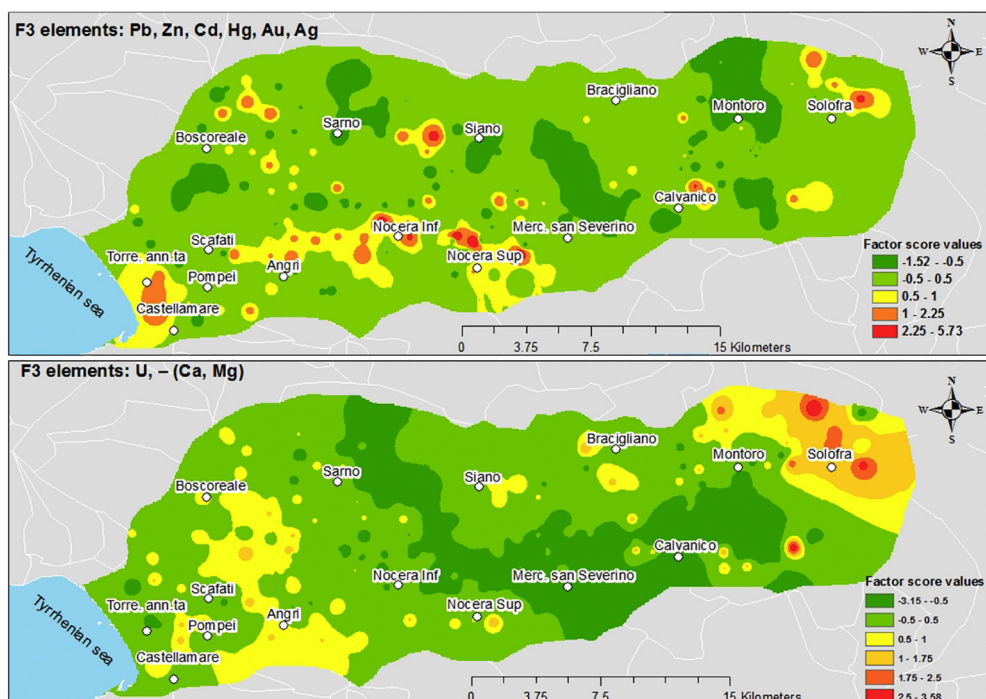


Fig. 8. Factor scores association map in soils of the survey area; (upper) Factor scores map F3 (Pb, Zn, Cd, Hg, Au, Ag) using lognormal data, and (lower) factor scores map F3 (U, - (Ca, Mg)) using clr transformed data.

- The distribution of Ni and Cr (relating to the Ni–Cr association) in the Sarno Basin generally occurs in soil located along the banks of the Sarno River. This cannot be attributed to the underlying geology and is most likely a result of the discharge of wastewater into the river from the tanneries of Solofra. This association is clearer in interpolated factor score maps, using the clr-transformed data.
- The influence of both geogenic and anthropogenic activities on the elevated elemental concentrations in the Sarno River Basin can be

clearly identified using the interpolated factor association maps, and results of the clr-biplot using clr-transformed coordinates and after amalgamating variables.

- In addition, the correlation between variables highlighted by factor analysis methods, is a useful tool in the interpretation and determination of the spatial distribution of the patterns, taking into account both their geogenic and anthropogenic sources. Factor association score map NRD F2 in conjunction with CLR F3 (the Ca–Mg

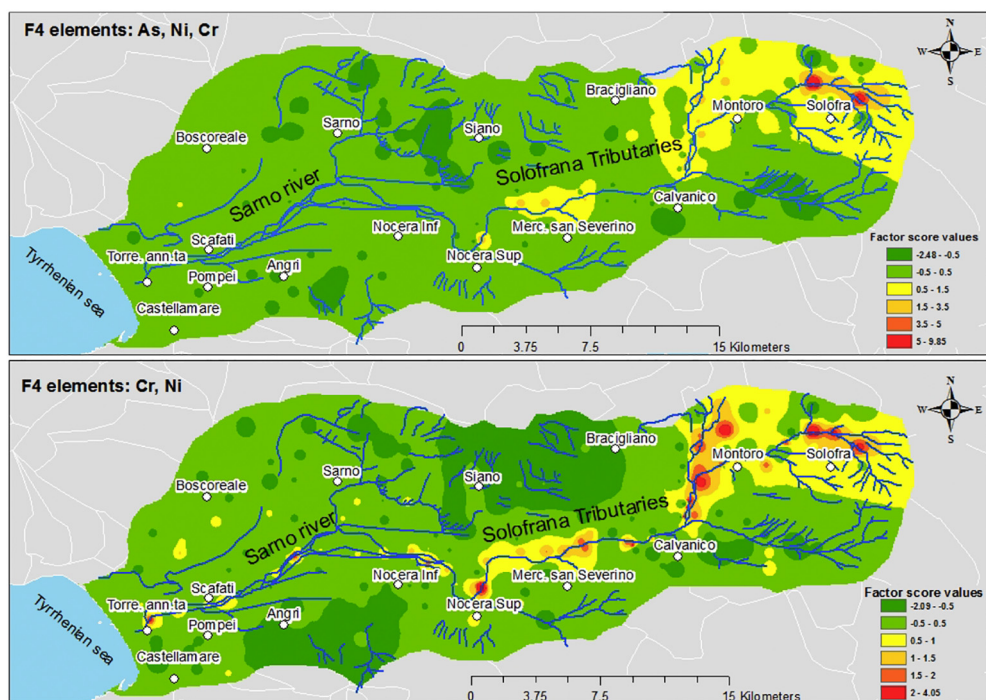


Fig. 9. Interpolated factor score map showing the elemental association in soils of the Sarno River Basin. (Upper) Factor scores map F4 (As, Ni, Cr) using lognormal data, and (lower) factor scores map F4 (Cr, Ni) using clr transformed data.

association), and factor score map NRD F1 and CLR F1 (the Na, K, P, Rb, Ba, B, Zr, Li, As, Th, and La associations), clearly show high factor scores reflecting the dominant underlying lithology occurring throughout the Sarno River Basin.

Future work could be undertaken to further distinguish the proportionality of the mixed behavior seen in certain elements (i.e., Pb, Zn, Cd, and Hg), by applying other multivariate statistical tools (such as robust principal component and discrimination analysis) and/or advanced environmental geochemistry methods such as the use of biomarkers. This may continue to highlight the environmental conditions in the Sarno Basin, and provide a framework in which to work to improve the quality of life for resident population.

Acknowledgements

The present work has been developed in the framework of the project “Development of green technologies for production of BIOchemicals and their use in preparation and industrial application of POLimeric materials from agricultural biomasses cultivated in a sustainable way in Campania region (BioPoliS)” (PON03PE_00107_1), funded in the frame of Operative National Program Research and Competitiveness 2007–2013 D. D. Prot. N. 713/Ric. 29/10/2010.

References

- Adamo, P., Zampella, M., Gianfreda, L., Renella, G., Rutigliano, F.A., Terribile, F., 2006. Impact of river overflowing on trace element contamination of volcanic soils in south Italy: part I. Trace element speciation in relation to soil properties. *Environ. Pollut.* 144, 308–316.
- Aitchison, J., 1986. *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman and Hall Ltd (reprinted 2003 with additional material by The Blackburn Press), London, UK (1986).
- Aitchison, J., Egozcue, J.J., 2005. Compositional data analysis: where are we and where should we be heading? *Math. Geol.* 07, 829–850.
- Aitchison, J., Greenacre, M., 2002. Biplot of compositional data. *Appl. Stat.* 51 (4), 375–392.
- Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J.A., Pawłowsky-Glahn, V., 2000. Logratio analysis and compositional distance. *Math. Geol.* 32 (3), 271–275.
- Albanese, S., Iavazzo, P., Adamo, P., Lima, A., De Vivo, B., 2013. Assessment of the environmental conditions of the Sarno River Basin (south Italy): a stream sediment approach. *Environ. Geochem. Health* 35, 283–297.
- Aral, M., 2010. *Environmental Modeling and Health Risk Analysis (Acts/Risk)*. Springer, Dordrecht Heidelberg London New York. <http://dx.doi.org/10.1007/987-90-481-8608-2>.
- Arienzo, M., Adamo, P., Rosaria Bianco, M., Violante, P., 2001. Impact of land use and urban runoff on the contamination of the Sarno River Basin in southwestern Italy. *Water Air Soil Pollut.* 131, 349–366.
- Barceló-Vidal, C., Martín-Fernández, J.A., 2016. The mathematics of compositional analysis. *Austrian J. Stat.* 45 (4), 57–71.
- Chayes, F., 1960. On correlation between variables of constant sum. *J. Geophys. Res.* 65 (12), 4185–4193.
- Cheng, Q., 2008. Modeling local scaling properties for multiscale mapping. *Vadose Zone J.* 7, 525–532.
- Cicchella, D., Giaccio, L., Lima, A., Albanese, S., Cosenza, A., Civitillo, D., De Vivo, B., 2014. Assessment of the topsoil heavy metals pollution in the Sarno River Basin, south Italy. *Environ. Earth Sci.* 71, 5129–5143.
- Comas-Cufí, M., Thió-Henestrosa, S., 2011. CoDaPack 2.0: a stand-alone, multi-platform compositional software. In: Egozcue, J.J., Tolosana-Delgado, R., Ortego, M.I. (Eds.), *CoDaWork'11: 4th International Workshop on Compositional Data Analysis*. SantFeliu de Guíxols.
- De Pippo, T., Donadio, C., Guida, M., et al., 2006. *Environ. Sci. Pollut. Res. Int.* 13, 184. <http://dx.doi.org/10.1065/espr2005.08.287>.
- DEFRA (Department for Environment, Food and Rural Affairs), 2011. *Guidelines for Environmental Risk Assessment and Management: Green Leaves III*. Defra and the Collaborative Centre of Excellence in Understanding and Managing Natural and Environmental Risks. Cranfield University, UK.
- Eckart, G., Young, G., 1936. The approximation of one matrix by another of lower rank. *Psychometrika* 1 (3), 211–218.
- Egozcue, J.J., Pawłowsky-Glahn, V., 2006. Simplicial geometry for compositional data. In: Buccianti, A.A., Mateu-Figueras, G., Pawłowsky-Glahn, V. (Eds.), *Compositional Data Analysis in the Geosciences: From Theory to Practice*. Geological Society, London, pp. 145–159.
- Egozcue, J.J., Pawłowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. *Math. Geol.* 35 (3), 279–300.
- Filzmoser, P., Hron, K., 2011. Robust statistical analysis. In: Pawłowsky-Glahn, V., Buccianti, A. (Eds.), *Compositional Data Analysis. Theory and Applications*. John Wiley & Sons, Chichester, UK, pp. 59–72.
- Gabriel, K.R., 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58 (3), 453–467.
- Health Canada, 2010a. *Federal Contaminated Site Risk Assessment in Canada. Part I: Guidance on Human Health Preliminary Quantitative Risk Assessment (PQRA)*. Version 2.0. Health Canada, Ottawa.
- Health Canada, 2010b. *Federal Contaminated Site Risk Assessment in Canada, Part II: Health Canada Toxicological Reference Values (TRVs) and Chemical-Specific Factors*. Version 2.0. Health Canada, Ottawa.
- ISTAT, 2013. *Resident Population on 1 January 2013 Available Via Dialog*. <http://www.istat.it/it/campania/dati>, Accessed date: 25 April 2016.
- Kempton, R.A., 1984. The use of biplots in interpreting variety by environment interactions. 103 (01), 123–135.
- Mateu-Figueras, G., Daunis i Estadella, J., 2008. *Compositional Amalgamations and Balances: A Critical Approach*. CODAWORK'08. Available via Dialog. <http://hdl.handle.net/10256/738>.
- Miesch Programs, 1990. G-RFAC. Grand Junction, Colorado, USA.
- Otero, N., Tolosana-Delgado, R., Solera, A., Pawłowsky-Glahn, V., Canals, A., 2005. Relative vs. absolute statistical analysis of compositions: a comparative study of surface waters of a Mediterranean river. *Water Res.* 39, 1404–1414.
- Palarea-Albaladejo, J., Martín-Fernández, J.A., Soto, J.A., 2012. Dealing with distances and transformations for fuzzy C-means clustering of compositional data. *J. Classif.* 29 (2), 144–169.
- Pawłowsky-Glahn, V., Egozcue, J.J., Lovell, D., 2015a. Tools for compositional data with a total. *Stat. Model.* 15, 175–190.
- Pawłowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R., 2015b. *Modelling and Analysis of Compositional Data*. Wiley, Chichester.
- Pearson, K., 1897. Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc. Roy. Soc. Lond.* 60, 489–498.
- Plant, J.A., Klaver, G., Locutura, J., Salminen, R., Vrana, K., Fordyce, F.M., 1996. *Forum of European Geological Surveys (FOREGS) Geochemistry Task Group 1994–1996*. British Geological Survey (BGS) Technical Report WP/95/14.
- Reimann, C., De Caritat, P., 2005. Distinguishing between natural and anthropogenic sources for elements in the environment: regional geochemical surveys versus enrichment factors. *Sci. Total Environ.* 337, 91–107.
- Rossini, O.S., Fernández, E.A.J., 2007. Monitoring of heavy metals in topsoil, atmospheric particles and plant leaves to identify possible contamination sources. *Microchem. J.* 86, 131–139.
- Salminen, R., Tarvainen, T., Demetriades, A., Duris, M., Fordyce, F.M., Gregorauskiene, V., et al., 1998. *FOREGS Geochemical Mapping Field Manual*. Espoo: Guide 47. Geological Survey of Finland.
- Tariq, R.S., Shah, M.H., Shaheen, N., Khaliq, K., Manzoor, M., Jaffar, S., 2005. Multivariate analysis of trace metal levels in tannery effluents in relation to soil and water: a case study from Peshawar, Pakistan. *J. Environ. Manag.* 79, 20–29.
- Thiombane, M., Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Cavaliere, M., De Vivo, B., 2017. Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model. *J. Geochem. Explor.* 189, 85–99.

Paper 5

Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model

Matar Thiombane, Daniela Zuzolo, Domenico Cicchella, Stefano Albanese, Annamaria Lima, Marco Cavaliere, Benedetto De Vivo

Journal of Geochemical Exploration, Volume 189, June 2018, Pages 85-99



Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model

Matar Thiombane^{a,1}, Daniela Zuzolo^{b,1}, Domenico Cicchella^{b,*}, Stefano Albanese^a, Annamaria Lima^a, Marco Cavaliere^a, Benedetto De Vivo^a

^a Department of Earth, Environment and Resources Sciences, University of Napoli "Federico II", Complesso Universitario di Monte Sant'Angelo, Via Cintia snc, 80125 Napoli, Italy

^b Department of Science and Technology, University of Sannio, Via Port'Arsa 11, 82100 Benevento, Italy

ARTICLE INFO

Keywords:

Campania region
Geochemical anomalies
Log-ratio transformed data
Compositional biplot
C-A fractal model
GeoDAS

ABSTRACT

Campania Region (southern Italy) is characterized by several elemental soil anomalies both geogenic and anthropogenic. Parts of these anomalies occur into the World Heritage Territory known as "National Park of the Cilento and Vallo di Diano", where this follow-up study has been carried out. In this paper, a methodology based on compositional data analysis (CoDA) and factor score C-A fractal model was applied on geochemical data from the above study in order to identify geochemical signatures associated anomalies. Eighty-one top soil samples were collected over an area of 98 km², and analyzed by ICP-MS after aqua regia digestion. Frequency based method (edaplots, classical and robust compositional biplot) and frequency space-method (factor score maps) were applied to visualize the correlation between variables and their main features into the survey areas. The different geochemical patterns were distinguished by a multivariate analysis combined with Concentration-Area (C-A) fractal method. Results show that geochemical data should be transformed under a compositional perspective to avoid artefacts, prior to statistical computations. Indeed, *ilr*-transformed data show a distinct bimodal distribution for several elements. This type of distribution appears masked considering raw and lognormal data. A "robustification" of the variables dataset permits to found a more clear relationship between variables. Factor score maps based on *ilr*-transformed variables and C-A plot cut off threshold displayed different geochemical patterns. In our survey area soil alteration phenomena could mask the nature of parental rock. The factor scores maps highlight an antithetic behaviour of many elements. Elements characterized by elevated geochemical mobility, such as Ca and Mg, are negatively correlated with Chemical Alteration Index values. Instead, Sn, Th, Be, Al are characterized by elevated geochemical stability; for this reason they are found where high-weathered soils occurs. The presence of elements such as Co, Cu Fe, Ni, Cr, Zn, K, Mn is mainly controlled by terrigenous flysch deposition. As, Pb (exceeding the CSC) and Sb association mainly occur in correspondence of urban areas and where traffic jams are frequent.

1. Introduction

In the last two decades, geochemical mapping has been recognized as a relevant method both in mineral exploration and environmental geochemistry. The compositional nature of geochemical data has long been neglected. These types of data are compositional because they belong to a closed system (each element is a part of a whole) and they should be treated as such to avoid spurious correlation (Pawlowsky-Glahn and Buccianti, 2011a). To treat the compositional geochemical

data in univariate or multivariate analysis, we must move the D -part simplex S^D , to real space, $R^D - 1$ or R^D , using log-ratio transformations (additive log-ratio (alr), centered log-ratio (clr) and isometric log-ratio (ilr)) that assume the sample in Aitchison geometry which can properly reconcile the correlation between elements (Aitchison, 1986; Egozcue et al., 2003; Filzmoser and Hron, 2008; Filzmoser et al., 2009a; Filzmoser et al., 2009b; Carranza, 2011; Zuo et al., 2013). Different methods have been used for processing geochemical data and generating spatial distribution interpolated maps from point data. "They

* Corresponding author.

E-mail address: domenico.cicchella@unisannio.it (D. Cicchella).

¹ Equal contributors.

belong to frequency-based and frequency-space-based methods” (Zuo et al., 2013). The frequency-based methods are based on the analysis of data frequency distributions using univariate and multivariate tools (Filzmoser et al., 2009a). Among the frequency-space-based methods, fractal and multifractal analysis (Cheng et al., 1994; Mandelbrot, 1983) have been demonstrated to be a powerful tool for identifying geochemical anomalies (e.g., Cheng, 1999a, 2007; Cheng et al., 2000c, 2010; Lima et al., 2003; Cicchella et al., 2015; Zuo, 2011; Zuo et al., 2015) and separate them from background values. The anomaly analysis is one of the most discussed issues because it has been helpful to improve the quality of geochemical interpretation of natural and anthropogenic phenomena (Cicchella et al., 2005). These exploratory methods of geochemical data have been widespread applied to soil geochemical data analysis. In this study we propose an exploratory compositional data analysis and concentration-area (C-A) fractal model applied on factor scores of the element associations obtained by Factor Analysis to detect the origin of multi-element soil anomalies (geogenic, anthropogenic or mixed). The latter procedure was applied on data from the soil follow-up geochemical survey in the Cilento - Vallo di Diano area, stemming from a regional survey study covering the entire Campania region, southern Italy (Buccianti et al., 2015; De Vivo et al., 2016). Campania region is characterized by several elemental soil anomalies of both geogenic and anthropogenic sources (Albanese et al., 2007; De Vivo et al., 2016; Lima et al., 2003; Zuzolo et al., 2016). Parts of these anomalies occur into the World Heritage Territory known as “National Park of the Cilento and Vallo di Diano”. Previous regional-scale studies on stream sediments (Albanese et al., 2007) and in soils (De Vivo et al., 2016) showed interesting different potentially toxic elements anomalies (PTEs). In particular, for soils in different sites occur values exceeding the contamination thresholds (CSC) established by Italian legislation for soils (Legislative Decree 152/2006, n.d.). The attention on the Cilento - Vallo di Diano area was also individuated by Minolfi et al. (2016) study, which indicated the presence of several PTEs exceeding the CSC. A geochemical follow-up survey was carried out on the aforementioned area, in order to assess the distribution pattern of geochemical anomalies and to distinguish the natural sources (geogenic) from the anthropogenic ones. The main objectives of this study were: (1) to illustrate the importance of compositional log-transformations in geochemical data for anomaly features assessment; (2) to delineate their main sources by multivariate analysis and GIS-based approach; (3) to use robust compositional data evaluation to assess geogenic or anthropogenic conditions of multi-element associations when the classical biplot is sensitive to outliers; (4) to prove the usefulness of factor score maps using C-A fractal model for processing geochemical data and to distinguish different geochemical distribution patterns.

The geochemical data analysis was performed taking account both of the compositional and their fractal distribution nature.

2. Univariate analysis and compositional biplot

To better visualize the element distributions and their main different sources into the survey area, univariate and multivariate computations were displayed through edaplots and compositional biplot, respectively. Edaplot (Exploratory Data Analysis Plot) is a combination of histogram, density plot, on dimensional scattergram and a boxplot and is considered one of the best statistical tools to display the data distribution (Reimann et al., 2008). The histogram is able to display unimodal (symmetric distribution or skewed) or multimodal distribution. The one-dimensional scattergram itself draws up the data along a straight line while the density plot traces a variation of the histogram that uses kernel smoothing to plot values; the picks and the form of the shape allow the interpretation of the data distribution.

Compositional biplot is a powerful statistical tool which displays both samples (observations) and the variables of a data matrix in terms of the resulting scores and loading (Gabriel, 1971). Thus the scores

represent the structure of the compositional data hold into a Euclidian space based on variance and covariance matrix; moreover they display the association structure of the dataset where the rays are define from the center of the plot and the length of vectors are proportional to the amount of explained variance (communality) of the variables they represent. The interpretation of the graphic depends on the loading (rays) structures and in more details the approximate links between rays and samples, the distances between vertexes and their directions (Otero et al., 2005). The utility of compositional biplots is better described in several papers (Pison et al., 2003; Maronna et al., 2006; Filzmoser and Hron, 2008; Filzmoser et al., 2009b; Otero et al., 2005; Hron et al., 2010) related to their use on compositional data elaboration.

Classical compositional biplot (CCB) and Robust Compositional Biplot (RCB) are used to compute the correlation between variables using compositional raw data and log-transformed data. However the biplot of raw data is substantially influenced by occurrence of outliers and the fact that it misleads the compositional nature of the data matrix which can affect the principal components in results interpretation (Aitchison, 1986; Filzmoser et al., 2009a, 2009b). For these reasons log-transformed data are recommended to use in multivariate analysis, in addition to the robust version of compositional biplot. Taking into account the singularity of the clr transformation (Aitchison, 1986), compositional data should be computed in orthonormal basis such as ilr transformed data, and back-transformed to the clr space for further interpretation. This back transformation allows to conserve the linear relationship between clr and ilr coordinates (Egozcue et al., 2003), applying the minimum covariance determinant (MCD) estimator (Rousseeuw and Van Driessen, 1999). The MCD looks for a subset h out of n observations of the initial data with the lowest determinant of sample covariance matrix (Filzmoser and Hron, 2008) and allow obtaining variables into normal distribution which describe better the correlation or covariance of the data avoiding the outlier artefacts.

3. Concentration-Area fractal model (C-A plot in factor scores maps)

Factor analysis was used to investigate the complex multivariate relationships among variables, which are not properly put on display by simple correlation analysis. Factor analysis is a multivariate statistical technique for identifying different groups of chemical elements with approximately similar geochemical patterns. It extracts the most important information from the data on the concept of communality (for each variable, communality is defined as the common variance explained by the factors). In order to facilitate the interpretation of results, varimax rotation was applied, since it is an orthogonal rotation that minimizes the number of variables that have high loadings on each factor, simplifying the transformed data matrix and assisting interpretation. Factor analysis simplifies the interpretation because the different factors obtained can be related to the different lithologies, surface enrichment phenomena, geochemical barriers or anthropogenic contamination.

In practice factor analysis reduces a large number to a much smaller number of variables called factors, characterized by elemental association factor scores. The distribution of elemental association factor scores was investigated. Geochemical point data can be processed by different methods to generate spatial distribution interpolated maps. The frequency-space-based methods have been demonstrated to be a powerful tool for identifying geochemical anomalies in several studies (e.g., Lima et al., 2003; Cicchella et al., 2005, 2015; Cheng, 1999a, 2000a; Cheng et al., 2010; Zuo et al., 2015; Zuo and Wang, 2016).

Cheng et al. (1994) proposed a concentration-area (C-A) fractal model for geochemical anomaly separation. From a multifractal point of view, the extreme (anomalous) values may follow a fractal distribution rather than a normal or lognormal law. The concentration-area (C-A Plot) fractal method separates anomalies from background on the basis of concentration values, as well as the spatial and geometrical

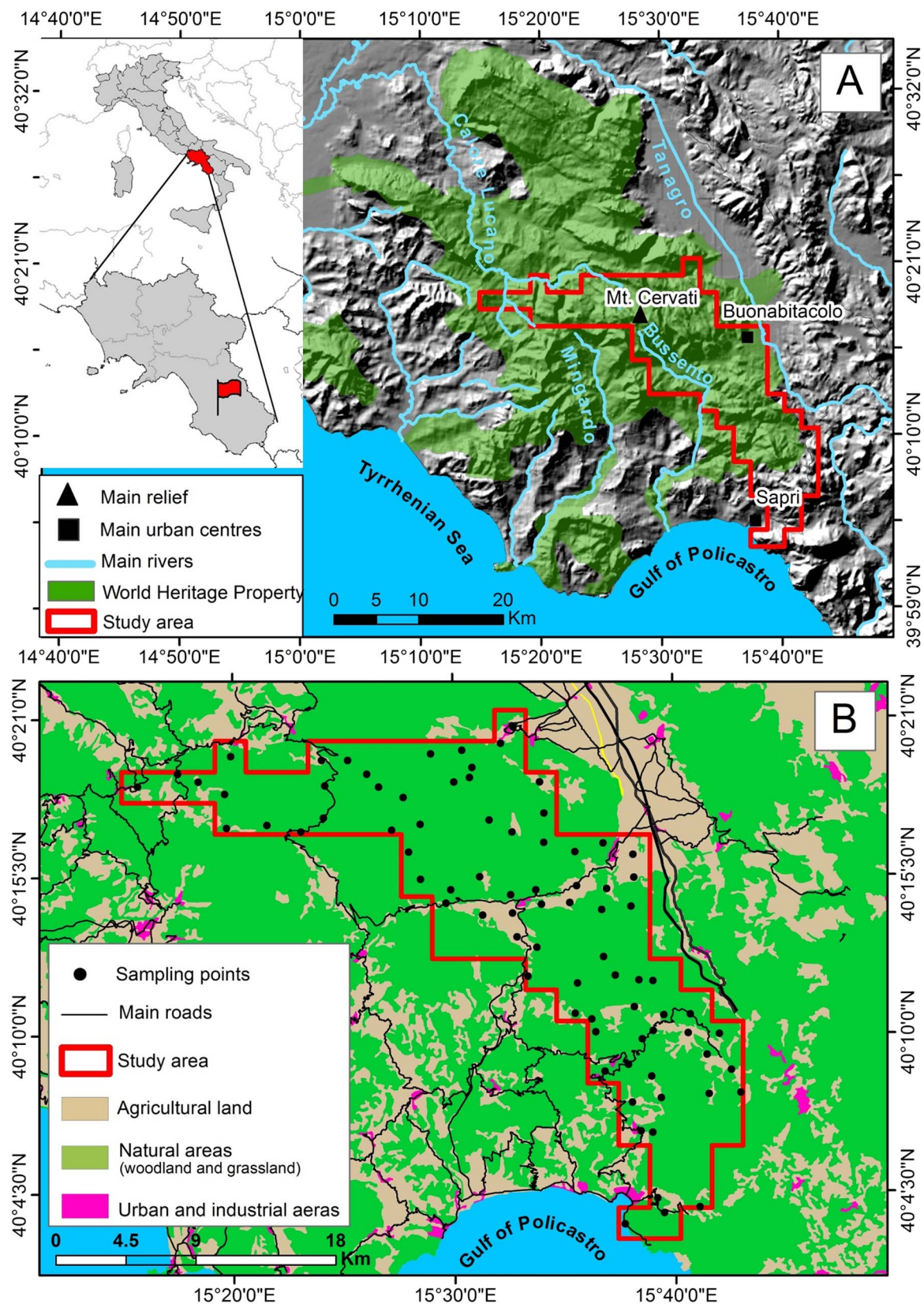


Fig. 1. Introductory map: A Localization and main geomorphological features of the study area. B Land use, road map and sampling point location of survey area (Cilento, Southern Italy).

properties of geochemical patterns. Point data can be interpolated by multifractal-IDW (MIDW) interpolation based on GeoDAS software (GeoData Analysis System for Mineral Exploration and Environmental Assessment) (Cheng, 1999b, 2000a; Lima et al., 2003). A C-A plot on log-log paper can be used to establish power-law relationships between

the area $A(\geq s)$ with the concentration values greater than s and the concentration value s itself. On a log-log paper a fractal curve is calculated and breakpoints (change in slope of the curve) represent different statistical populations. Several straight-line segments can be fitted in correspondence of breakpoints to provide a set of cut-off values

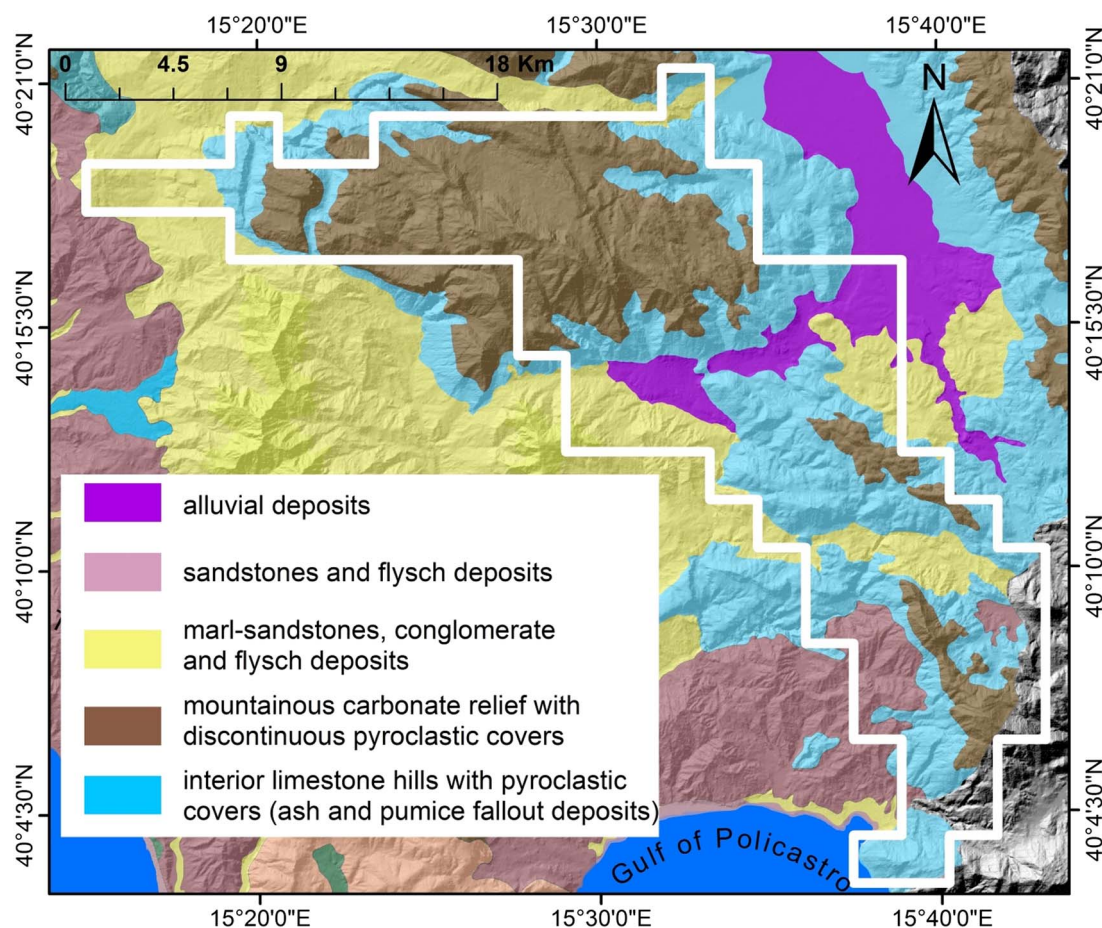


Fig. 2. Land system map of the study area (Cilento, southern Italy). Modified from Di Gennaro et al., 2002.

for subdividing the concentration scale into discrete classes. These discretizations define different geochemical population interpretable as specific features of the area.

In this case, C-A fractal method was applied on the elemental association factor scores to highlight anomalies of different geochemical processes into the survey area.

4. Materials and methods

4.1. Survey area

The study area covers an area of 98 km² in south-west of Italy and settles in the southern part of the Campania Region. Most of the survey area falls within a World Heritage Territory known as “National Park of the Cilento and Vallo di Diano”. The geomorphology of the area is predominantly mountainous and hilly constituted by Cervati Mountain (the highest mountain of Campania region with 1899 m high). Most of the survey area is drained by Bussento and Mingardo rivers (Fig. 1A) and is mainly characterized by the presence of natural outlines such as grasslands and persistent woodlands (Fig. 1B). Heterogeneous agricultural areas are mainly concentrated in less steep areas (Cilento plain) with extensive orchards, vineyards, and chestnuts, greenhouse horticulture, floriculture, and crop production throughout and around the main urban zones (Buonabitacolo, Sapri). The survey area is crossed by different trunk roads and the A3 highway which skirts the area and connects the main urban centers. Urban environments are limited, discontinuous and concentrated mostly along the coastal side.

The survey area is part of the Southern Apennines fold and thrust belt formed as a result of the convergence between the Apulian and the European plates since the late Cretaceous (Vitale et al., 2010). In the

soils of Italy, as well as in our survey area, the nature of the parental rock can often be masked by the effects of climate and vegetation and alteration degree, that play a fundamental role in determining soil properties and development (Costantini and Dazzi, 2013). For this reason and in line with the purposes of this study, we propose a land system map in Fig. 2 (mod. from Di Gennaro et al., 2002). This map highlights areas with typical “patterns” of lithology and morphology; the occurring patterns are governed by climate, geology and landform which interact over time to influence the distribution of soils and vegetation. It is possible to observe the predominance of mountainous carbonate reliefs with discontinuous pyroclastic coverings and widespread limestones hills controlled by ash and pumice fallout deposits. The Quaternary volcanic activity of Vesuvius (100 km NW), Phlegrean Fields (Campi Flegrei) (120 km NW), Roccamonfina (200 km NW), Mt. Vulture (located east of the Apennine mountain range, about 100 km NE) and Aeolian islands volcanic arc has exerted an influence on the formation of soils in the study area, giving a widespread pyroclastic footprint (Buccianti et al., 2015; Scheib et al., 2014). Moreover, marl-conglomerate, marl-sandstone and alluvial deposits occur in the study area.

4.2. Sampling and sample preparation

During summer 2016, 81 topsoil samples were collected from the study area (98 km²) at an average sampling density of approximately one sample per 1.2 km². About 1.5 kg soil samples was collected at a depth between 5 and 15 cm under the ground surface after the vegetation cover removal, following the protocol according to the FOREGS sampling procedures (Plant et al., 1996; Salminen et al., 1998). Each sample is made up from a composite soil material taken from five points

Table 1

Detection limit, accuracy and precision of the applied analytical method (RPD = relative percent difference; %DL = percentage of samples with concentrations below the detection limit).

Elements	Unit	Detection limit (DL)	% < DL	Precision ^a (%) RPD)	Accuracy ^b (%)
Al	%	0.01	0	3.1	3.4
Ca	%	0.01	0	3.3	4.6
Fe	%	0.01	0	2.2	3.4
K	%	0.01	0	1.9	2.4
Mg	%	0.01	0	1.9	3.0
Na	%	0	0	1.4	6.0
P	%	0	0	5.9	7.4
S	%	0.02	6	5.2	1.2
Ti	%	0	2	3.4	9.2
As	mg/kg	0.1	0	7.8	4.8
B	mg/kg	1	0	2.0	9.0
Ba	mg/kg	0.5	0	10.0	5.3
Be	mg/kg	0.1	0	21.2	18.0
Bi	mg/kg	0.02	0	6.7	10.1
Cd	mg/kg	0.01	0	7.4	4.5
Co	mg/kg	0.1	0	2.1	1.0
Cr	mg/kg	0.5	0	2.4	5.6
Cu	mg/kg	0.01	0	7.2	4.3
Ga	mg/kg	0.1	0	5.8	4.4
La	mg/kg	0.5	0	5.7	10.1
Li	mg/kg	0.1	0	2.1	2.2
Mn	mg/kg	1	0	1.7	6.0
Mo	mg/kg	0.01	0	3.1	4.0
Ni	mg/kg	0.1	0	3.6	2.1
Pb	mg/kg	0.01	0	3.5	5.0
Sb	mg/kg	0.02	0	7.4	7.3
Se	mg/kg	0.1	2	15.2	7.3
Sn	mg/kg	0.1	0	4.2	12.5
Sr	mg/kg	0.5	0	5.1	10.0
Th	mg/kg	0.1	0	3.5	8.4
Tl	mg/kg	0.02	0	1.1	3.0
U	mg/kg	0.1	0	13.6	11.9
V	mg/kg	2	0	4.5	3.1
W	mg/kg	0.1	14	3.0	2.0
Zn	mg/kg	0.1	0	5.7	8.1
Ag	µg/kg	2	0	4.5	24.9
Au	µg/kg	0.2	2	3.6	17.7
Hg	µg/kg	5	0	7.0	4.2

^a Precision was calculated as relative percentage difference (%RPD) using the formula: %RPD = $[(SV - DV) / (SV + DV) / 2] \times 100$, where SV = the original sample value, DV = the duplicate sample value.

^b The laboratory accuracy error was determined using the formula: Accuracy error = $[(X - TV) / TV] \times 100$, where X = laboratory's analysis result for the performance sample (standard) and TV = true value of the performance sample (standard).

over 5 m square and in every 20 sampling sites a duplicate sample was collected in the same cell to allow the blind control of the cell sampling variability combined with analytical variability. At each sampling spatial coordinates, topography, local geology, type and main properties of soils, land use, and any additional detail related to anthropic activities in the surroundings were recorded. The 81 samples were dried with infra-red lamps at a temperature below 35 °C, pulverized in a ceramic mortar and then sieved to retain the < 2 mm fraction. The pulps were stored in small plastic bags containing at least 30 g of samples for chemical analysis.

4.3. Chemical analysis

Analyses were carried out by Mineral Laboratories Bureau Veritas (Vancouver, Canada). Each sample was digested in aqua regia solution and analyzed by ICPMS (inductively coupled plasma-mass spectrometry) for 53 elements (Ag, Al, As, Au, B, Ba, Be, Bi, Ca, Cd, Ce, Co, Cr, Cs, Cu, Fe, Ga, Ge, Hf, Hg, In, K, La, Li, Mg, Mn, Mo, Na, Nb, Ni, P, Pb, Pd, Pt, Rb, Re, S, Sb, Sc, Se, Sn, Sr, Te, Th, Ti, Tl, U, V, W, Y, Zn and Zr), as required by Italian environmental Law (D. Lgs. 152/2006). The digestion by modified aqua regia is considered to give “pseudo total”

concentrations of metals (Albanese et al., 2007). The quality of data was assessed by estimation of accuracy and precision (Table 1). Precision of the analysis was calculated using in-house replicates, while accuracy was determined using laboratory's in-house reference material.

4.4. Data processing and factor score maps

In this paper, the production of univariate (Eadplots) and multivariate (Compositional biplot) graphics was implemented by the free and open source statistical software R, one the most used statistical tools to provide descriptive statistics and graphics to advanced methods (e.g. robust compositional tools). Two mains open-source packages for R were used for compositional data analysis: “Compositions” (Van Den Boogaart et al., 2011) and “Robcompositions” (Templ et al., 2011).

The univariate computations (Eadplots) were displayed using the raw data, lognormal data and ilr log transformed data of As, Pb and Co. The ilr log transformation was applied on data taking into account the compositional vectors of n parts partitioned into groups of parts representing a certain affinity (Egozcue et al., 2003). “It is based on the choice of an orthonormal basis (in the well-known Euclidean sense) in the hyperplane formed by the clr transformation.” (Filzmoser and Hron, 2009c). The orthonormal basis were built defining non-overlapping groups through a sequential binary partition (SBP) of the whole composition (Filzmoser and Hron, 2009c; Pawlowsky-Glahn and Egozcue, 2011b). The coordinates which represent an element of the simplex in the orthonormal basis defined by an SBP represents the balances (Egozcue and Pawlowsky-Glahn, 2005). This computation is shown in Table 2 which contains 18 different elements or parts in the simplex S^{18} , and this information can be expressed within 17 dimensions in R^{17} that will form the balances. In this study, the affinity throughout variables or “partition of elements” was assessed by applying a hierarchical cluster analysis (Fig. 3) which displays links between parts. For the multivariate computations, such as robust biplot and factor analysis, it is usefulness to use balances (Z_1, Z_2 , etc....) loadings and scores in plot because is not straightforward to interpret in sense of composition. Indeed, the variables can be retraced applying a clr back transformation of the balances (Z_1, Z_2 , etc....) to their corresponding variables (As, Cu, etc....) in the clr space. This back transformation doesn't affect the covariance matrix between variables due to the fact that it exists a linear relationship between variables ilr and clr transformations (Egozcue et al., 2003) and ilr transformation expresses the clr coordinates in a orthonormal basis (Filzmoser et al., 2009a, 2009b).

To better visualize the correlation between variables and display the main behaviour of these variables into the survey areas, 18 elements were considered (Table 3). The reduction of the number of elements is based on 3 main criteria: 1) removal of elements with > 40% of values below the limit of detection (LOD); 2) choose arbitrary mostly two representative elements (e.g., Rare Earth Elements; REE), which are geochemically congruent; 3) choose elements with a communality of extraction higher than 0.5 (50%) and/or a common variances < 0.5 (e.g. Reimann et al., 2002).

Factor score distribution maps were performed by means of GeoDAS (Cheng, 1999b, 2000a, 2000b) using Multifactorial Inverse Distance Weighted (MIDW) script as an interpolation method (Cheng, 1999b) and ArcGIS software. Factor scores have been classified using the concentration–area fractal method (C-A). This method allows images to be subdivided into components for symbolizing distinct image zones representing specific features on the ground. Considering that the factor scores values present negative and zero values which are not “log transformable”, was applied a min-max normalization by scaling the original data within a specified range of features (e.g., ranging from 1 to 100). Furthermore, min-max normalization performs a linear transformation on the original data without changing the geometrical structure of the simplex (Han and Kamber, 2001).

Suppose that X and Y are the minimum and the maximum values for

Table 2

Sequential binary partitions and resulting balances of 18 elements of the survey areas. Elements are grouped referring to the hierarchical cluster which takes into account their affinity.

Balance	Ca	Mg	Zn	P	K	Th	Sn	Al	Be	As	Pb	Sb	Ni	Cr	Fe	Cu	Co	Mn
Z ₁	+	–																
Z ₂			+	–														
Z ₃					+	+	+	–	–									
Z ₄					+	+	–											
Z ₅					+	–												
Z ₆								+	–									
Z ₇										+	+	–						
Z ₈										+	–							
Z ₉													+	+	+	–	–	–
Z ₁₀													+	+	–			
Z ₁₁													+	–				
Z ₁₂																+	–	
Z ₁₃																	+	–
Z ₁₄	+	+	–	–														
Z ₁₅	+	+	+	+	–	–	–	–	–									
Z ₁₆	+	+	+	+	+	+	+	+	+	–	–	–						
Z ₁₇	+	+	+	+	+	+	+	+	+	+	+	+	–	–	–	–	–	–

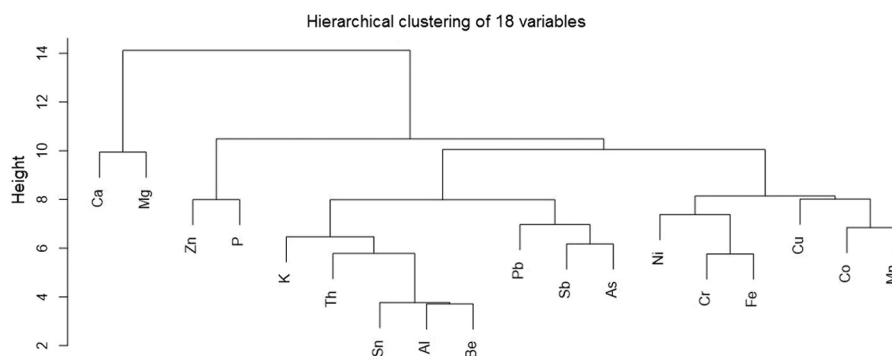


Fig. 3. Hierarchical cluster of 18 variables shown as a dendrogram which displays the relationships between elements and their affinity; this cluster is plotted using the average linkage between variables.

initial dataset S . Min-max normalization maps a value for v of S to v' in the range $[\mu, \partial]$ by computing:

$$v' = \frac{(V - X)}{(Y - X)} \times (\partial - \mu) + \mu$$

where μ and ∂ relate the new minimum and maximum of the specified new range, successively.

Specifically, thresholds were determined by a C-A plot, in which the

vertical axis represents cumulative pixel areas $A(p)$, with element concentration values greater than p , and the horizontal axis the normalized factor scores values (p). Initial factor scores values will be obtained by computing a back-transformed of the normalized factor scores data. Breaks between straight-line segments and corresponding values of p were used to reclassify maps of elemental association factor scores maps.

Table 3

Statistical parameters for 18 elements of 81 topsoil samples of the survey area.

Element	Unit	Number of samples	Min	Max	Mean	Median	RMS	Std. deviation	Skewness	Kurtosis
Al	mg/kg	81	6300.00	69,700.00	36,541.98	31,900.00	40,009.77	16,394.63	0.42	– 0.92
As	mg/kg	81	2.50	65.10	16.09	14.70	19.52	11.12	1.74	4.57
Be	mg/kg	81	0.20	7.00	3.25	3.10	3.72	1.83	0.36	– 1.09
Ca	mg/kg	81	2600.00	288,700.00	37,007.41	16,500.00	61,818.42	49,825.88	2.81	9.02
Co	mg/kg	81	2.90	27.50	14.29	14.20	15.07	4.83	0.48	0.58
Cr	mg/kg	81	11.50	106.70	41.88	40.20	45.29	17.33	1.38	3.13
Cu	mg/kg	81	8.80	151.23	38.34	35.34	42.26	17.90	3.59	19.25
Fe	mg/kg	81	6200.00	59,700.00	32,785.19	33,100.00	34,114.33	9488.44	– 0.08	0.17
K	mg/kg	81	800.00	5400.00	3439.51	3600.00	3603.19	1080.36	– 0.04	– 0.83
Mg	mg/kg	81	2000.00	63,300.00	7439.51	5300.00	10,957.92	8095.61	4.79	27.47
Mn	mg/kg	81	118.00	2239.00	959.63	916.00	1028.33	371.86	0.67	1.22
Ni	mg/kg	81	9.40	81.50	37.43	37.40	39.45	12.53	0.54	1.42
P	mg/kg	81	100.00	2710.00	835.68	790.00	960.17	475.77	1.54	3.29
Pb	mg/kg	81	9.21	165.10	38.37	40.83	43.49	20.59	2.75	15.72
Sb	mg/kg	81	0.17	3.66	1.04	1.01	1.22	0.64	1.33	3.11
Sn	mg/kg	81	0.50	4.90	2.32	2.20	2.54	1.04	0.44	– 0.68
Th	mg/kg	81	1.10	23.70	8.79	6.70	10.24	5.28	1.01	0.21
Zn	mg/kg	81	30.80	392.40	103.21	102.50	111.04	41.20	4.20	28.32

4.5. Chemical alteration degree

Weathering indices are conventionally used to quantify the extent of weathering within a sample soil and help to explain processes occurring during pedogenesis. In our survey area, as previously described, soil alteration phenomena could mask the nature of parental rock. The extent of alteration was quantified using Chemical Index of Alteration (CIA). This index was proposed by Nesbitt and Young (1982) and has been extensively applied as a proxy to examine weathering. The CIA index is calculated as:

$$CIA = \frac{Al_2O_3}{Al_2O_3 + CaO + Na_2O + K_2O} * 100$$

The calculation of the CIA is a tool to measure the weathering phenomena of the geological feature of a survey area; it's well established certain range of values (percentage) to describe the level of weathering. In consequence, the raw concentration values are likely the more suitable values for this computation because the log transformed data (coordinates) present

null and negative values which are not useful rapport of ratio (on denominator) computation. During weathering the proportion of alumina to alkalis would typically increase in the weathered product and that is a good measure of the degree of weathering. The CIA values for the soil samples varied in the range 2.85–88.90 (average = 62.20), showing a high variability of the degree of weathering. In general, CIA values under 50 are characteristic of unweathered soils; 100 correspond to the optimum weathered value. Consequently, soil are considered weathered when clay minerals (e.g. montmorillonites) dominate, generally in the range of 75–85 (Nesbitt and Young, 1982; Price and Velbel, 2003).

5. Results and discussions

The chemical characterization of the survey area soils shows several PTEs (As, Be, Cd, Co, Cu, Pb, Sn, Tl, V and Zn) exceeding the contamination thresholds (CSC) established by Italian legislation for soils (Legislative Decree 152/2006, n.d.). Many of these elemental exceeds are very interesting due their spatial distribution (Fig. 4). For example

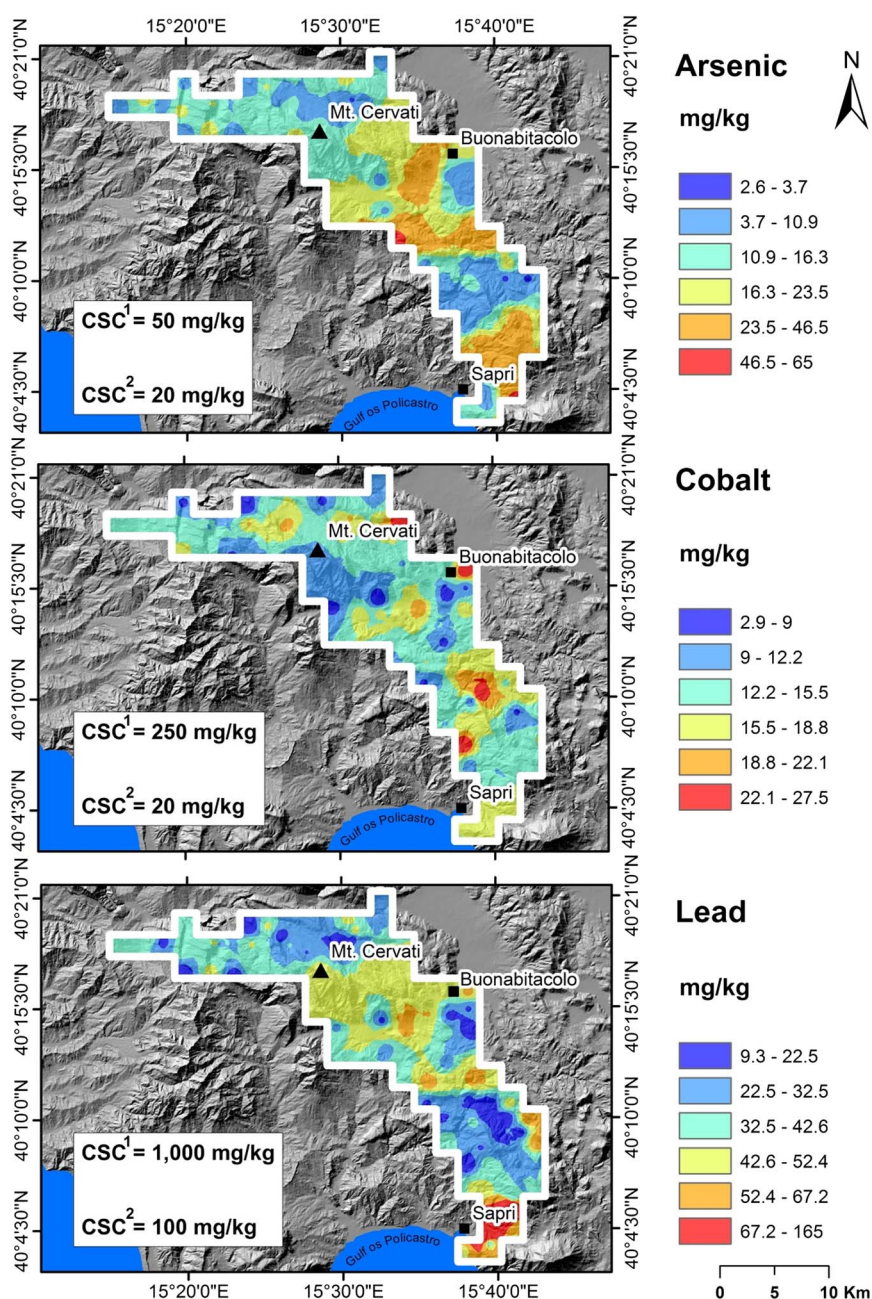


Fig. 4. Geochemical map of As, Co and Pb, compiled using multifractal-IDW interpolation method (GeoDas). Pixel values have been reclassified by fractal concentration–area (C–A) plot based on the frequency distribution of pixel values. CSC¹ = contamination thresholds established by Italian legislation for industrial/commercial use of soils (Legislative Decree 152/2006). CSC² = contamination thresholds established by Italian legislation for residential use of soils (Legislative Decree 152/2006).

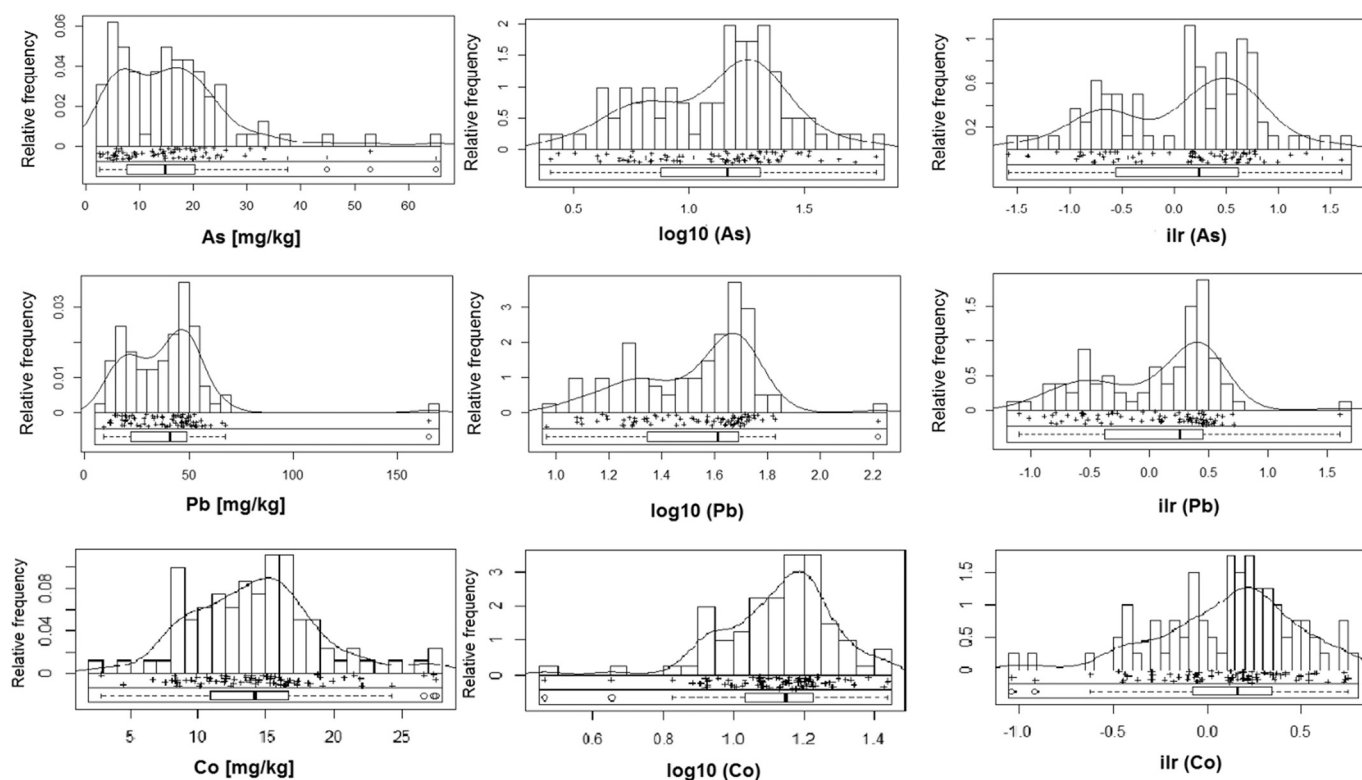


Fig. 5. Edaplots (combination of histogram, density trace, one-dimensional scattergram and Boxplot in just one display) of the raw data, log-transformed data, and ilr transformed data.

Be and As concentration are extremely high. Based on previous studies at national scale (Cicchella et al., 2015), it is clear that such high concentrations represent natural background values of these elements exceeding the Italian statutory limit for soil on the whole territory of southern and central Italy. The Pb spatial distribution seems to be controlled mostly by anthropogenic contributes, showing the highest concentration values in correspondence of the main urban centers (e.g., Sapri). Also Co shows concentrations over CSC for residential/recreational use (about 10% of analyzed samples > 20 mg/kg). This type of data analysis does not take into account their compositional nature. Locally different geochemical contributes could exist. A data analysis CODA-based could be helpful in the interpretation of the phenomena determining the soil geochemistry of the investigated area.

Fig. 5 displays the data distributions of As, Pb and Co into the survey area through an edaplot (combination of histogram, density trace, scattergram and boxplot). this figure shows three different representations of the data: the original concentrations data (left plots), the log-normal-transformed data (middle plots), and the ilr-transformed data (right plots). The graphical data show that the variables do not follow a normal distribution and tend to display a bimodal distribution clearly fitting in the ilr-transformed data plots. As and Pb clearly show two populations in ilr-transformed plots. This observation is clearly displayed on the histogram combined with density traces and the one-dimensional scattergram; this might reveal the main activities associated to these elements into the survey area. The boxplots show several data points in upper inner fence far away from the center of the whisker plotting the raw data; these artefacts tend to disappear or to be highly reduced when log-transformed and ilr-transformed data are used. In addition, these data points are considered as outliers and should be identified and modeled to avoid incorrect results. It proves that compositional data transformation such as ilr transformation solves this problem and moves the composition sample space to the Euclidean Real space R^{-1} , which is failed by the classic statistical data transformation.

The compositional biplot (Fig. 6) based on principal component analysis displays the correlation and relationship between 18 analytical

variables in 81 sample points, and the first two principal components extracted. The principal components are presented in a compositional biplot using raw data (Fig. 6, left plot) and ilr coordinates clr back-transformed (Fig. 6, right plot).

The total variance of initial raw data biplot explains 54.31%, where the first principal component (PC1) accounts 39.80% and the second principal component (PC2) accounts 14.51% (Fig. 6, left plot), whereas the robust biplot based on ilr coordinates clr back transformed produces much more suitable results (Fig. 6, right plot) with PC1 explaining 46.06% and PC2 explaining 19.36% of the compositional variability.

The classical biplot shows clearly three data (variable) elemental associations, whose observations (samples) are widely displayed into the graphic, emphasizing the presence of many outliers or artefacts correlated with Ca and Mg variables. These outliers are widely reduced with the robust compositional biplot. Nevertheless, the classical biplot displays three features of element associations which may be related to the typical lithological background of the survey area:

The Ca and Mg association (A1) presents both variables with high communalities (the length of the ray) with Ca more representative (87.02% of variability) and variables overlapping one to another. This is expected due to the fact that they belong to the same source. These two elements characterize the Mt. Cervati limestones (Trias-Lias), belonging to the Alburno-Cervati-Pollino stratigraphic unity.

Nickel, Fe, Mn, Cr, P, Zn, Cu and Co (A2) form clearly an association, their vertexes laying closer together and also because they tend toward the same direction. Fe and Mn are the most representative elements in A2 association due to the fact that the length their rays (variability) dominate in the group. Zn and P display short vectors which likely characterized poorly presented and seems to be only partially correlated to the others element of A2 association. However, Fe and Mn prove their predominance in the geochemical processes of this group and are likely related to the occurrence of adsorption phenomena. A2 association might be related to the adsorption and coprecipitation effects operated by Fe and Mn oxides and hydroxides occurring mostly in the sedimentary deposits such as marl-sandstone,

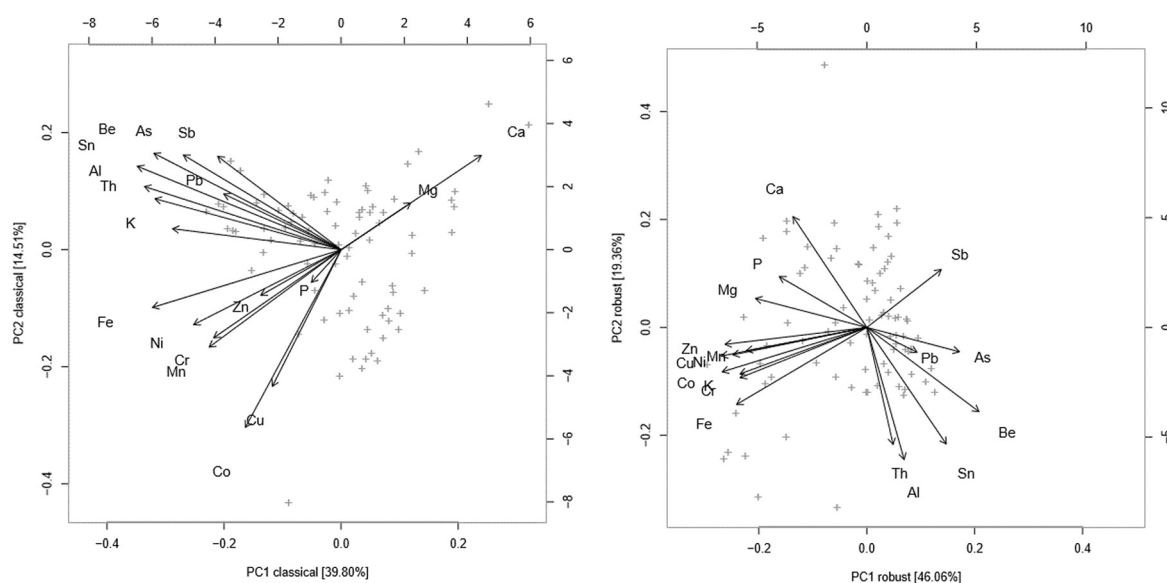


Fig. 6. Biplots for first and second principal components of factor analysis for raw data (classical, left plot) and ilr coordinates clr back-transformed (robust, right plot) of the survey area (Cilento, Southern Italy).

Table 4

Varimax-rotated factor (three-factor model) of isometric logratio clr back-transformed variables for 81 topsoil samples from the survey area; bold entries: loading values over |0.50|.

Variables	Rotated component matrix			Communalities
	1	2	3	
Ca	0.07	−0.80	0.04	0.65
Mg	0.35	−0.51	0.35	0.52
Zn	0.83	−0.18	0.16	0.75
P	0.34	−0.50	0.59	0.68
K	0.70	0.04	0.50	0.69
Th	0.25	0.77	−0.16	0.69
Sn	0.17	0.82	−0.18	0.73
Al	0.24	0.75	0.50	0.82
Be	−0.45	0.64	0.23	0.67
As	−0.38	0.29	−0.64	0.64
Pb	0.22	0.07	−0.55	0.51
Sb	−0.12	−0.25	−0.86	0.82
Ni	0.87	−0.04	0.01	0.75
Cr	0.84	0.19	0.01	0.74
Fe	0.9	0.32	−0.09	0.93
Cu	0.91	−0.04	0.03	0.82
Co	0.95	0.06	0.06	0.91
Mn	0.69	−0.08	0.16	0.55
Eigenvalues	6.71	3.74	2.16	
Total variance in %	37.28	20.8	12.06	
Cum. of total variance	37.28	58.08	70.14	

conglomerate and silico-clastic flysch deposits outcropping in the survey areas.

The third group is represented by Th, Sn, Be, K, Sb, Al, As, Pb and Sb association (A3). Al, Be and Th are elements with the highest variability in this association (A3) due to the fact that the lengths of their rays are larger. Indeed, it's likely related to the fact that these elements are mostly immobile during weathering phenomena of the parental rocks and mostly remain in the residual fraction of soil. Furthermore, this elemental association reflects the influence of pyroclastic deposits from different eruptions of nearby volcanoes (Roccamonfina, Vesuvius, Phlegrean Fields - De Vivo et al., 2010; Buccianti et al., 2015; Mt. Vulture and Aeolian islands - Peccherillo, 2005; Scheib et al., 2014) covering the carbonate (mostly limestone) of the study areas.

The robust PCA biplot “opened” data, displaying better the element

associations and reducing the outliers which might influence the results. The length of the rays is linked to the variability of the ilr transformed data back transformed in clr space, and not to the variables themselves. The robust biplot separates the A3 group of the classical biplot in two real associations which emphasize the chemical structures of the lithological background and the main human activities: Al, Th, Sn and Be association might be correlated to the pyroclastic deposits, whereas, Sb, Pb and As group might be related to the main human activities (vehicle emissions). It appears interesting the display of As ray in the two compositional biplots. In fact the As vertex merges inside the A3 association (controlled by pyroclastic deposits) of the classical biplot. Instead, for the robust biplot it separates two main groups and is located between these two latter main groups: Th, Be and Sn group related clearly to pyroclastics and Pb and Sb group to anthropogenic sources. This might reflect a mixed contribution: both geogenic and anthropogenic. In addition, the robust biplot shows P positioned between Ca and Mg revealing that phosphate is sorbed by calcite as commonly occurs in carbonate-rich alkaline soils.

Basically the compositional biplot was able to distinguish particular groups of elements, whereas the factor score maps indicate the incidence of identified associations in each sampled point allowing us to interpret possible correlations to natural peculiarities of the area (e.g. lithologies) and human activities, hence making it possible to discriminate geogenic vs anthropogenic sources of different elements. Moreover, C-A plot computation considers the compositional nature of geochemical data and their fractal distribution.

Factor analysis was performed on the 18 selected elements (K, Ni, Co, Cr, Fe, Cu, Zn, Sn, Al, Th, Be, Mn, P, Sb, Pb, Ca, Mg and As), using their ilr transformed data clr back transformed A three-factor model, accounting for 70.1% of total data variability, has been chosen and the varimax-rotated factors are reported in Table 4. The KMO (Kaiser-Meyer-Olkin) test is over 0.75, indicating a good adequacy of the proposed model. Elements with loadings over the value of 0.50 have been considered to describe the main composition of each factor. The elements of the associations of the three-factor model, sorted in ascending loading, are:

- F1: Co, Cu Fe, Ni, Cr, Zn, K, Mn
- F2: Sn, Th, Al, Be, - (Ca, Mg, P)
- F3: P, Al, K - (Sb, As, Pb)

The factor score distribution maps were classified using the concentration–area (C-A) fractal method implemented in GeoDAS software.

The F1 elemental association (Co, Cu Fe, Ni, Cr, Zn, K, Mn) accounts

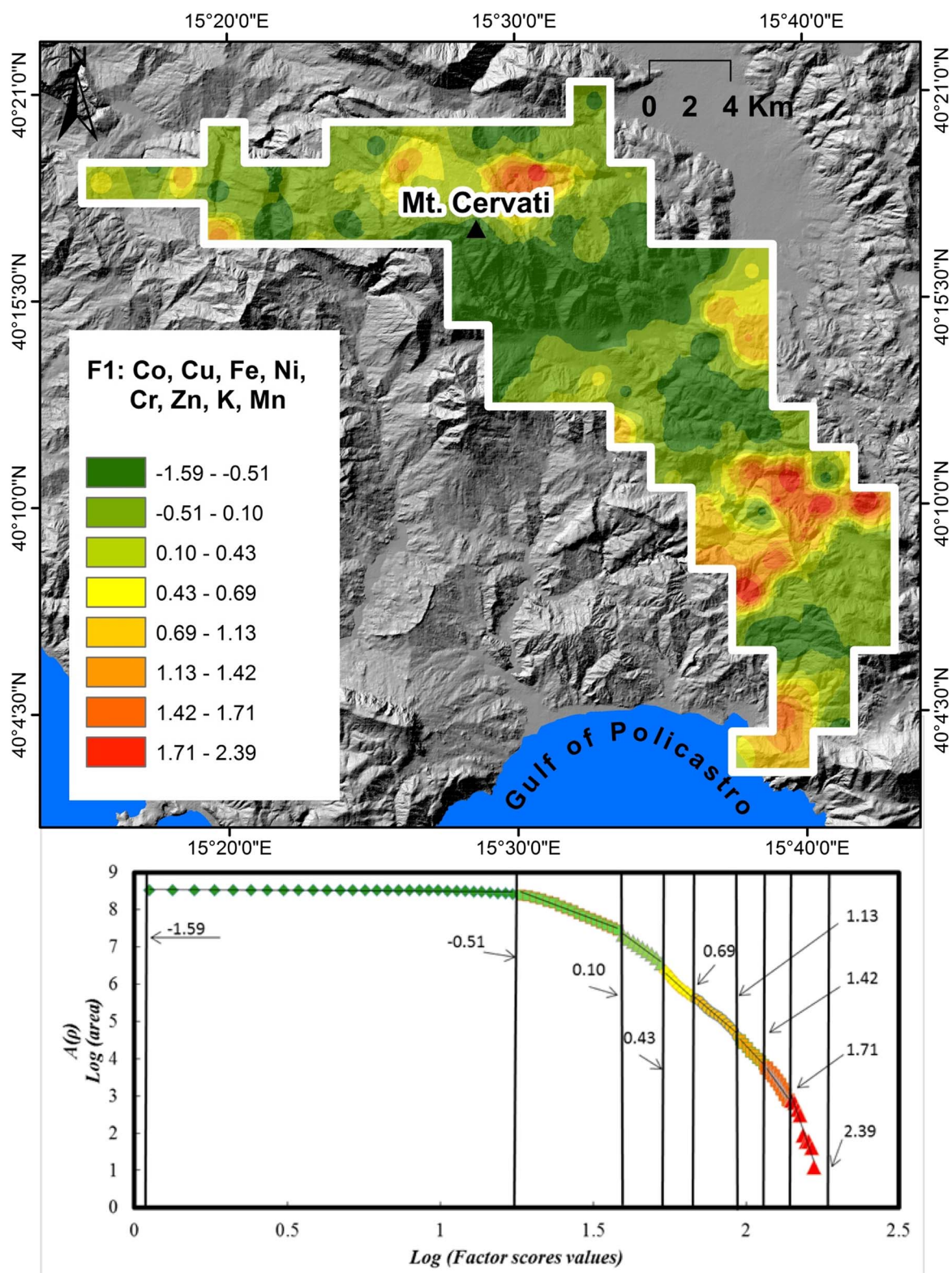


Fig. 7. Map showing the spatial distribution of the Factor 1 of the logratio transformed data; maps are classified through C-A plot (shown below).

37.3% of total variance. The factor scores interpolated map (Fig. 7) highlights that the highest values (ranging from 1.71 to 2.39) coincide exactly with the silico-clastic deposits dominated by flysch series. Fine-size solids with charged substrate such as clay minerals (e.g., smectite/montmorillonite) play an important role in the sorption of trace elements. Moreover, enrichment of trace elements highlighted by this association reflects adsorption and co-precipitation by Fe and Mn oxy-

hydroxides in oxidizing environment. The highest F1 factor score values were found in correspondence of the highest Mn concentrations (up to 1648 mg/kg).

The F2 elemental association [Sn, Th, Al, Be, - (Ca, Mg, P)] accounts a total variability of 20.8% and distinguishes the two groups of elements antithetically correlated (Sn, Th, Al, Be group versus Ca, Mg and P association). Highest factor score values (ranging from 1.74 to 2.21)

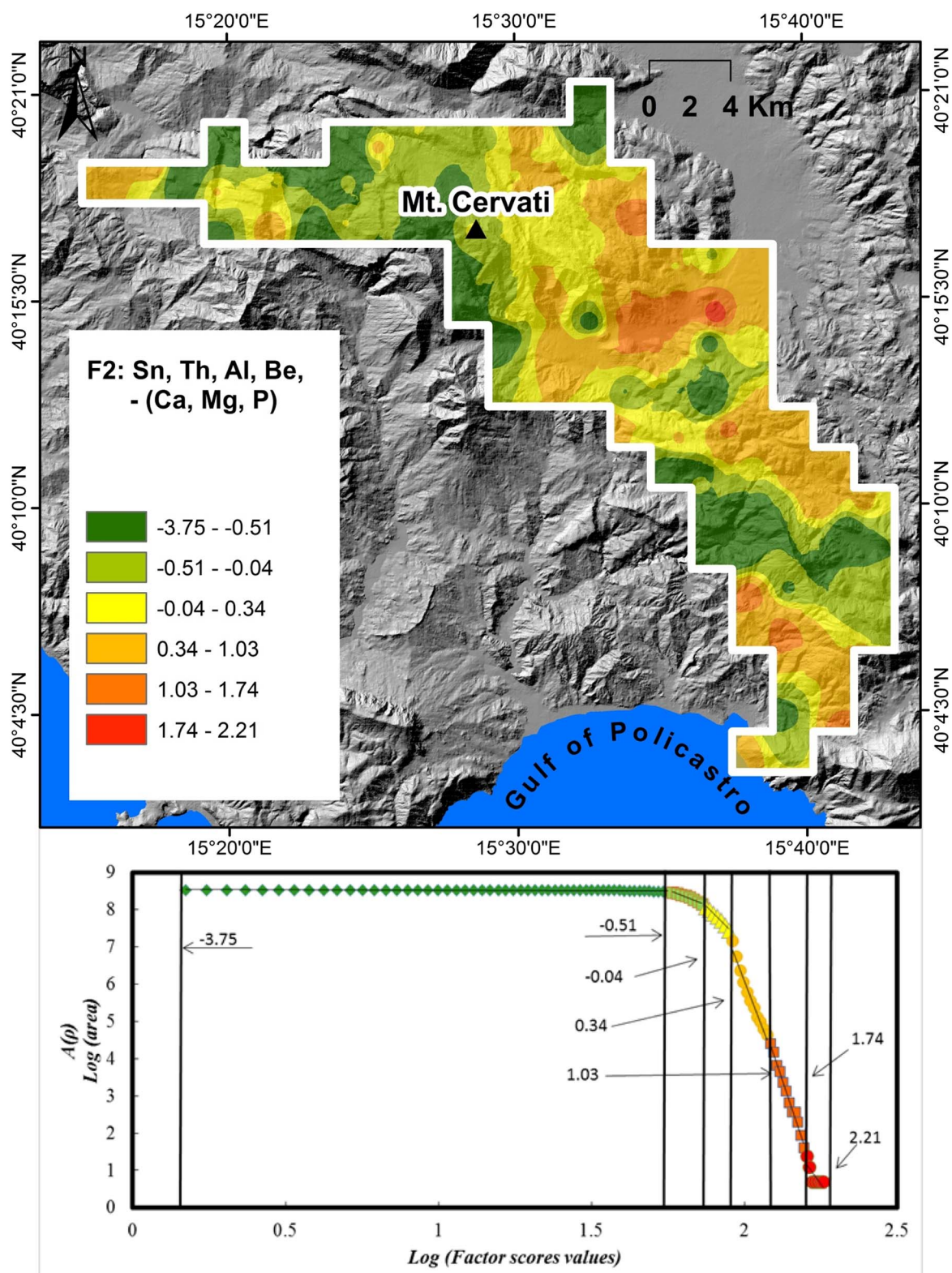


Fig. 8. Map showing the spatial distribution of the Factor 2 of the logratio transformed data; maps are classified through C-A plot (shown below).

of this association are located on the slope of the Mt. Cervati (Fig. 8). F2 maps clearly shows that Sn, Th, Be, Al are strongly controlled by presence of widespread pyroclastics linked to Quaternary volcanic activity of Vesuvius, Phlegrean Fields, Roccamonfina, Mt. Vulture and Aeolian islands (Peccerillo, 2005; De Vivo et al., 2010; Buccianti et al., 2015; Scheib et al., 2014) while Mg-Ca-P association reflects both the dominant carbonate lithology (limestones) of the Mt. Cervati and the

carbonate-rich alkaline soils where phosphate is commonly sorbed by calcite.

It is interesting to remark that F2 spatial distribution is well explained by the CAI of analyzed soils (Fig. 9). During weathering the proportion of alumina to alkalis would typically increase in the weathered product. Elements characterized by elevated geochemical mobility, such as Ca and Mg, are negatively correlated with CIA values.

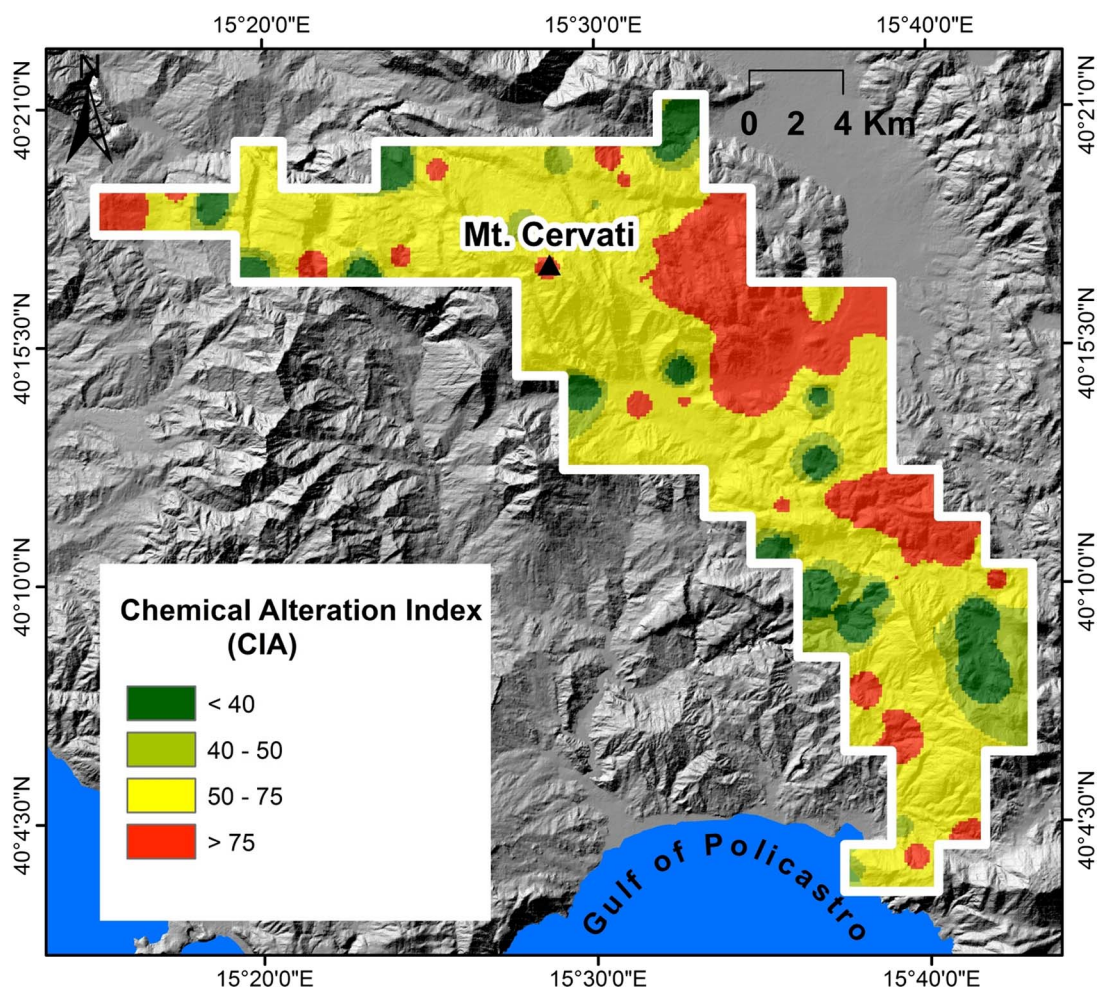


Fig. 9. Spatial distribution of Chemical Alteration Index (CIA). Values < 50 correspond to unweathered soils; > 50 weathered soils (Nesbitt and Young, 1982).

Accordingly Sn, Th, Be, Al, characterized by elevated geochemical stability, are found where high-weathered soils occurs. This evidence better clarify the antithetic behaviour of elements describing F2.

The F3 association P, Al, K - (Sb, As, Pb) displays 12.06% of total variance explained (Fig. 10). The highest factor score values (ranging from 0.70 to 2.44) indicative of a higher rate of P, Al, and K, are found mostly in the surrounding mountainous areas, at lower altitude, where soils are impacted by a number of natural and anthropogenic factors: weathering products from both limestone and pyroclastics and more densely populated areas. In any case, high factor scores occur in carbonate and clay-rich alkaline soils that readily incorporate K, limiting its mobility. Potassium may also reflect a contribution from limestone, as impure carbonate can contain up to 6% of K due to the occurrence of clays in non carbonate fraction (Wedepohl, 1978). Phosphorous, as mentioned earlier is sorbed by calcite.

The antithetic Sb, Pb and As elemental association is strong in correspondence of the most urbanized areas (Sapri, Buonabitacolo and Tortorella) and the main State roads intersections where traffic jams are frequent (Fig. 9). Although Sb does not exceed the contamination thresholds established by Italian legislation for soils (Legislative Decree 152/2006), Pb and As show values above the CSC nearby the urban area of Sapri. It is very interesting to note that the human footprint is clearly recognizable also in environmental contexts characterized by a low anthropogenic pressure (the study area mostly falls in a World Heritage Park).

6. Remarks and conclusion

The Cilento - Vallo di Diano area (Campania region, Southern Italy) is characterized by several PTEs exceeding the contamination thresholds (CSC) established by Italian legislation for soils (Legislative Decree 152/2006).

This study represents a geochemical follow-up of the mentioned area, in order to assess the distribution pattern and the behaviour of the geochemical anomalies, performing a methodology, which takes into account both compositionality and fractal distribution of geochemical data.

The ilr normalization moves the compositional data from the sample space to the Euclidean real space preserving the orthonormal basis. The frequency based methods (univariate) such as Edaplots of raw, log-transformed and ilr-transformed data were used to depict data distribution; the ilr-transformed variables show a clear bimodal distribution for As and Pb. This type of distribution tends to disappear when raw data and log-normal data are plotted. In fact data tend to be normalized when ilr-transformed data are used (Filzmoser et al., 2009b). Moreover, the main benefit of ilr-transformation, compared to raw data and log-normal data, is the enhancement of anomalous multi-element associations reflecting the geogenic, anthropogenic or mixed contribution affecting the geochemical dataset.

The classical biplot based on raw data shows the main geological feature of the study area identifying three lithological complexes such as:

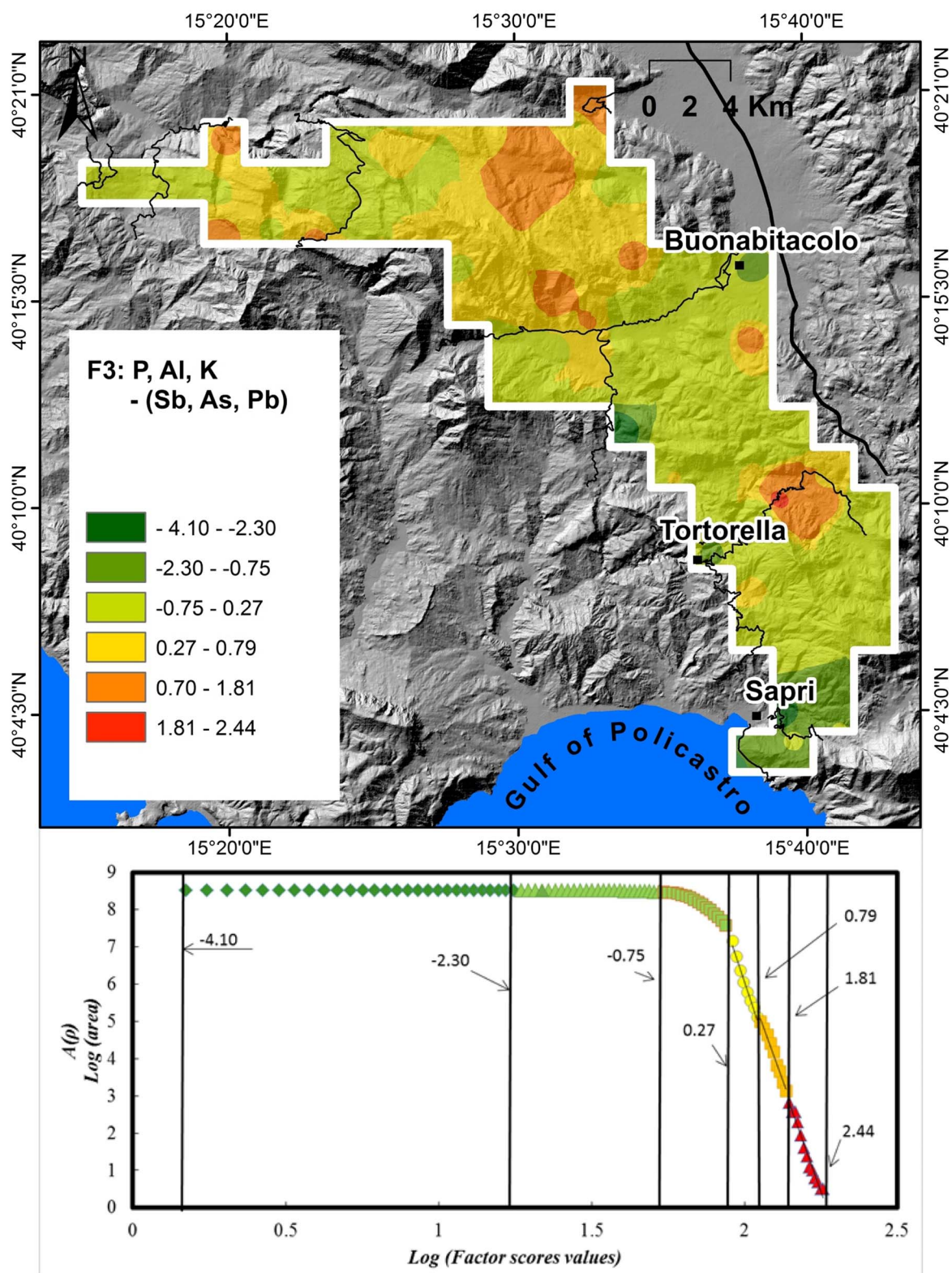


Fig. 10. Map showing the spatial distribution of the Factor 3 of the logratio transformed data; maps are classified through C-A plot (shown below).

- 1) a first association (Ca and Mg) reflecting limestones and dolostone occurrence of the Mt. Cervati;
- 2) a second association (Ni, Fe, Mn, Cr, Cu, Zn and Co) reflecting the coprecipitation effect of Mn and Fe oxides-hydroxides in flysch and arenaceous material of the survey area;
- 3) a third association (Th, Sn, and Be) related to the pyroclastic coverings.

The robust PCA based on a robust covariance estimator like the MCD, clearly shows the effect of the data opening. The variables in robust biplot (ilr coordinates clr back-transformed) are more evenly distributed across the component space, making easy the recognition of element associations characteristic of different geochemical (natural, anthropogenic or mixed) processes. This robustification clearly displays additional phenomena, which affect the natural features of the survey

area. In fact, Sb, Pb and As are well discriminated in the robust biplot and may be related to the influence of anthropogenic activities (vehicle emission).

The combination of C-A fractal model and factor analysis on log-transformed variables, allow us to recognize the geological features and the geochemical processes controlling the presence and distribution of chemical elements. In our survey area soil alteration phenomena could mask the nature of parental rock. The factor scores maps highlight an antithetic behaviour of many elements. Elements characterized by elevated geochemical mobility, such as Ca and Mg, are negatively correlated with CIA values. Instead, Sn, Th, Be, Al are characterized by elevated geochemical stability; for this reason they are found where high-weathered soils occurs. In the latter, surprisingly, also high F3 factor scores (with elements such as P and K) occur likely due to carbonate and clay-rich alkaline soils. The presence of elements such as Co, Cu Fe, Ni, Cr, Zn, K, Mn is mainly controlled by terrigenous flysch deposition. Arsenic, Pb (exceeding the CSC) and Sb association mainly occur in correspondence of urban areas and where traffic jams are frequent. It is very interesting to note how the human footprint (As, Sb Pb) is clearly recognizable also in environmental contexts characterized by low anthropogenic pressures (the study area mostly falls in a World Heritage Park).

According to our results and observations, this method demonstrate to be a useful tool to distinguish different processes controlling the elemental geochemical distribution in our study area, highlighting additional phenomena which would normally be masked.

Acknowledgments

This work was supported by grant from the Ministero dell'Università e della Ricerca Scientifica - Industrial Research Project "BioPoliS" PON03PE_00107_01, funded in the frame of Operative National Programme Research and Competitiveness 2007-2013, D. D. Prot. N. 713/Ric. 29/10/2010.

References

- Aitchison, J., 1986. *The Statistical Analysis of Compositional Data*. Chapman & Hall, London (416 pp).
- Albanese, S., De Vivo, B., Lima, A., Cicchella, D., 2007. Geochemical background and baseline values of toxic elements in stream sediments of Campania region (Italy). *J. Geochem. Explor.* 93, 21–34.
- Buccianti, A., Lima, A., Albanese, S., Cannatelli, C., Esposito, R., De Vivo, B., 2015. Exploring topsoil geochemistry from the CoDA (Compositional Data Analysis) perspective: the multi-element data archive of the Campania Region (Southern Italy). *J. Geochem. Explor.* 159, 302–316.
- Carranza, E.J.M., 2011. Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values. *J. Geochem. Explor.* 110, 167–185.
- Cheng, Q., 1999a. Spatial and scaling modelling for geochemical anomaly separation. *J. Geochem. Explor.* 65, 175–194.
- Cheng, Q., 1999b. In: Lippard, S.J., Naess, A., Sinding-Larsen, R. (Eds.), *Multifractal interpolation*. Proc. Ann. Conf. International Association for Mathematical Geology, Trondheim, Norway 6–11 August 1999. Vol. 1. Norwegian University of Science and Technology, Trondheim, pp. 245–250.
- Cheng, Q., 2000a. Interpolation by means of multifractal, kriging and moving average techniques. *GeoCanada 2000*. In: Proc. GAC/MAC Meeting, Calgary, AB, Canada [CDROM], 29 May–2 June 2000. Geol. Assoc. Can., St. John's, NF, Canada.
- Cheng, Q., 2000b. *GeoData Analysis System (GeoDAS) for Mineral Exploration: Unpublished User's Guide and Exercise Manual*. Material for the Training Workshop on GeoDAS held at York University.
- Cheng, Q., 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. *Ore Geol. Rev.* 32, 314–324.
- Cheng, Q., Agterberg, F.P., Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. *J. Geochem. Explor.* 51, 109–130.
- Cheng, Q., Xu, Y., Grunsky, E., 2000c. Integrated spatial and spectrum method for geochemical anomaly separation. *Nat. Resour. Res.* 9, 43–51.
- Cheng, Q., Xia, Q., Li, W., Zhang, S., Chen, Z., Zuo, R., Wang, W., 2010. Density/area power-law models for separating multi-scale anomalies of ore and toxic elements in stream sediments in Gejiu mineral district, Yunnan Province, China. *Biogeosciences* 7, 3019–3025.
- Cicchella, D., De Vivo, B., Lima, A., 2005. Background and baseline concentration values of elements harmful to human health in the volcanic soils of the metropolitan and provincial area of Napoli (Italy). *Geochem. Explor. Environ. Anal.* 5, 29–40.
- Cicchella, D., Giaccio, L., Dinelli, E., Albanese, S., Lima, A., Zuzolo, D., Valera, P., De Vivo, B., 2015. GEMAS: Spatial distribution of chemical elements in agricultural and grazing land soil of Italy. *J. Geochem. Explor.* 154, 129–142.
- Costantini, E.A.C., Dazzi, C., 2013. *The Soils of Italy*. Springer (354 pp.).
- De Vivo, B., Petrosino, P., Lima, A., Rolandi, G., Belkin, H.E., 2010. Research progress in volcanology in Neapolitan area, Southern Italy: a review and alternative views. *Mineral. Petrol.* 99, 1–28.
- De Vivo, B., Lima, A., Albanese, S., Cicchella, D., Rezza, C., Civitillo, D., Minolfi, G., Zuzolo, D., 2016. *Atlante geochimico-ambientale dei suoli della Campania* (Environmental Geochemical Atlas of Campania Soils). Aracne Editrice, Roma, pp. 364 (978-88-548-9744-1).
- Di Gennaro, A., Aronne, G., De Mascellis, R., Vingiani, S., Sarnataro, M., Abalsamo, P., Cona, F., Vitelli, L., Arpaia, G., 2002. I sistemi di terre della Campania. In: *Monografia e Carta*. 1. pp. 250.000. <http://hdl.handle.net/11588/179891>.
- Egozcue, J.J., Pawlowsky-Glahn, V., 2005. Groups of parts and their balances in compositional data analysis. *Math. Geol.* 37, 795.
- Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., et al., 2003. *Math. Geol.* 35, 279.
- Filzmoser, P., Hron, K., 2008. Outlier detection for compositional data using robust methods. *Math. Geosci.* 40 (3), 233–248.
- Filzmoser, P., Hron, K., 2009c. Correlation analysis for compositional data. *Math. Geosci.* 41 (8), 905–919.
- Filzmoser, P., Hron, K., Reimann, C., 2009a. Principal component analysis for compositional data with outliers. *Environmetrics* 20 (6), 621–632.
- Filzmoser, P., Hron, K., Reimann, C., 2009b. Univariate statistical analysis of environmental (compositional) data - problems and possibilities. *Sci. Total Environ.* 407, 6100–6108.
- Gabriel, K.R., 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58 (3), 453.
- Han, J., Kamber, M., 2001. *Data Mining: Concepts and Techniques*. Morgan-Kaufmann Academic Press, San Francisco.
- Hron, K., Templ, M., Filzmoser, P., 2010. Imputation of missing values for compositional data using classical and robust methods. *Comput. Stat. Data Anal.* 54 (12), 3095–3107. <http://dx.doi.org/10.1016/j.csda.2009.11.023>.
- Legislative Decree 152/2006 Decreto Legislativo 3 aprile 2006, n. 152, "Norme in materia ambientale". *Gazzetta Ufficiale* n. 88 14-4-2006, Suppl. Ord. n. 96. <http://www.camera.it/parlam/leggi/deleghe/06152dl.htm>.
- Lima, A., De Vivo, B., Cicchella, D., Cortini, M., Albanese, S., 2003. Multifractal IDW interpolation and fractal filtering method in environmental studies: an application on regional stream sediments of Campania Region (Italy). *Appl. Geochem.* 18 (12), 1853–1865. [http://dx.doi.org/10.1016/S0883-2927\(03\)00083-0](http://dx.doi.org/10.1016/S0883-2927(03)00083-0).
- Mandelbrot, B.B., 1983. *The Fractal Geometry of Nature*. Freeman, San Francisco, pp. 468.
- Maronna, R., Martin, R., Yohai, V., 2006. *Robust Statistics: Theory and Methods*. John Wileypp. 436 (ISBN: 978-0-470-01092-1).
- Minolfi, G., Albanese, S., Lima, A., Tarvainen, T., Fortelli, A., De Vivo, B., 2016. A regional approach to the environmental risk assessment - human health risk assessment case study in the Campania region. *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.jgexplo.2016.12.010>.
- Nesbitt, H.W., Young, G.M., 1982. Early Proterozoic climates and plate motions inferred from major element chemistry of lutites. *Nature* 299, 715–717.
- Otero, N., Tolosana-Delgado, R., Soler, A., Pawlowsky-Glahn, V., Canals, A., 2005. Relative vs. absolute statistical analysis of compositions: a comparative study of surface waters of a Mediterranean river. *Water Res.* 39, 1404–1414.
- Pawlowsky-Glahn, V., Buccianti, A., 2011a. *Compositional Data Analysis: Theory and Applications*. John Wiley & Sons.
- Pawlowsky-Glahn, V., Egozcue, J.J., 2011b. Exploring compositional data with the CoDa-dendrogram. *Austrian J. Stat.* 40 (1 & 2), 103–113.
- Peccherillo, A., 2005. Plio-quaternary volcanism in Italy. In: *Petrology, Geochemistry, Geodynamics*. Springer-Verlag, Berlin Heidelberg, pp. 365 (ISBN 978-3-540-29,092-6).
- Pison, G., Rousseeuw, P.J., Filzmoser, P., Croux, C., 2003. Robust factor analysis. *J. Multivar. Anal.* 84, 145–172.
- Plant, J.A., Klaver, G., Locutura, J., Salminen, R., Vrana, K., Fordyce, F.M., 1996. Forum of European Geological Surveys (FOREGS) geochemistry task group 1994–1996. In: *British Geological Survey (BGS) Technical Report WP/95/14*.
- Price, J., Velbel, M., 2003. Chemical weathering indices applied to weathering profiles developed on heterogeneous felsic metamorphic parent rocks. *Chem. Geol.* 202, 4397–4416. <http://dx.doi.org/10.1016/j.chemgeo.2002.11.001>.
- Reimann, C., Filzmoser, P., Garrett, R., 2002. Factor analysis applied to regional geochemical data: problems and possibilities. *Appl. Geochem.* 17 (3), 185–206.
- Reimann, C., Filzmoser, P., Garrett, R.G., Dutter, R., 2008. *Statistical data analysis explained*. In: *Applied Environmental Statistics with R*. Chichester. Wiley, pp. 362 (ISBN: 978-0-470-98581-6).
- Rousseeuw, P.J., Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41, 212–223.
- Salminen, R., Tarvainen, T., Demetriades, A., Duris, M., Fordyce, F.M., Gregorauskiene, V., Kahelin, H., et al., 1998. *FOREGS Geochemical Mapping Field Manual*. Geological Survey of Finland, Espoo (Guide 47). <http://www.gtk.fi/foregs/geochem/fieldmanan.pdf>.
- Scheib, A.J., Birke, M., Dinelli, E., The Gemas Project Team, 2014. Geochemical evidence of aeolian deposits in European soils. *Boreas* 43 (1), 175–192. <http://dx.doi.org/10.1111/bor.12029>.
- Templ, M., Hron, K., Filzmoser, P., 2011. *robCompositions: Robust Estimation for Compositional Data*. Manual and Package, Version 1.4.4.
- Van Den Boogaart, K.G., Tolosana-Delgado, R., Bren, R., 2011. *Compositions*.

- Compositional Data Analysis. R Package Version 1. pp. 10–12. Available at: <http://CRAN.R-project.org/package=compositions>.
- Vitale, S., Ciarcia, S., Mazzoli, S., Iannace, A., Torre, M., 2010. Structural analysis of the Internal Units of Cilento, Italy: new constraints of the Miocene tectonic evolution of the southern Apennine accretionary wedge. *Compt. Rendus Geosci.* 342, 475–482.
- Wedepohl, K.H. (Ed.), 1978. *Handbook of Geochemistry*. Springer-Verlag, Berlin-Heidelberg.
- Zuo, R., 2011. Decomposing of mixed pattern of arsenic using fractal model in Gangdese belt, Tibet, China. *Appl. Geochem.* 26, S271–S273.
- Zuo, R., Wang, J., 2016. Fractal/multifractal modeling of geochemical data: a review. *J. Geochem. Explor.* 164, 33–41.
- Zuo, R., Xia, Q., Wang, H., 2013. Compositional data analysis in the study of integrated geochemical anomalies associated with mineralization. *Appl. Geochem.* 28, 202–211.
- Zuo, R., Wang, J., Chen, G., Yang, M., 2015. Identification of weak anomalies: a multifractal perspective. *J. Geochem. Explor.* 148, 12–24.
- Zuzolo, D., Cicchella, D., Catani, V., Giaccio, L., Guagliardi, I., Esposito, L., De Vivo, B., 2016. Assessment of potentially harmful elements pollution in the Calore River basin (southern Italy). *Environ. Geochem. Health* 39 (3), 531–548.


Paper 6

Source patterns and characterisation of Polycyclic Aromatic hydrocarbons (PAHs) in urban and rural areas of Southern Italy

Matar Thiombane, Stefano Albanese, Marcello Di Bonito, Annamaria Lima, Roberto Rolandi, Shihua Qi, Benedetto De Vivo

Journal of environmental geochemistry and health, in press, available online 06th July 2018

Source patterns and contamination level of polycyclic aromatic hydrocarbons (PAHs) in urban and rural areas of Southern Italian soils

Matar Thiombane  · Stefano Albanese · Marcello Di Bonito · Annamaria Lima · Daniela Zuzolo · Roberto Rolandi · Shihua Qi · Benedetto De Vivo

Received: 10 January 2018 / Accepted: 29 June 2018
© Springer Nature B.V. 2018

Abstract Polycyclic aromatic hydrocarbons (PAHs) are a group of persistent organic pollutants. They have been identified as a type of carcinogenic substance and are relatively widespread in environment media such as air, water and soils, constituting a significant hazard for human health. In many parts of the world, PAHs are still found in high concentrations despite improved legislation and monitoring, and it is therefore vital defining their profiles, and assessing their potential sources. This study focused on a large region of the south of Italy, where concentration levels, profiles, possible sources and toxicity equivalent quantity (TEQ) level of sixteen PAHs were investigated. The survey included soils from five large regions of the south of Italy: 80 soil samples (0–20 cm top layer) from urban and rural locations were collected and analysed by gas chromatography–mass spectrometry.

Total PAHs and individual molecular compounds from the US Environmental Protection Agency priority pollutants list were identified and measured. Results showed that 16 PAHs varied significantly in urban and rural areas, and different regions presented discordant characteristics. Urban areas presented concentrations ranging from 7.62 to 755 ng g⁻¹ (mean = 84.85 ng g⁻¹), whilst rural areas presented ranges from 1.87 to 11,353 ng g⁻¹ (mean = 333 ng g⁻¹). Large urban areas, such as Rome, Naples and Palermo, exhibited high PAHs total concentration, but high values were also found in rural areas of Campania region. Different PAHs molecular ratios were used as diagnostic fingerprinting for source identification: LWMPAHs/HWMPAHs, Fluo/(Fluo + Pyr), BaA/(BaA + Chr), Ant/(Ant + Phe) and IcdP/(IcdP + BghiP). These ratios indicated that

M. Thiombane (✉) · S. Albanese · A. Lima · R. Rolandi
Department of Earth, Environment and Resources
Sciences (DiSTAR), University of Naples “Federico II”,
Complesso Universitario di Monte Sant’ Angelo, Via
Cintia snc, 80126 Naples, Italy
e-mail: thiombane.matar@unina.it

M. Di Bonito
School of Animal, Rural and Environmental Sciences,
Nottingham Trent University, Brackenhurst Campus,
Southwell NG25 0QF, UK

D. Zuzolo
Department of Science and Technology, University of
Sannio, Via dei Mulini 59/A, 82100 Benevento, Italy

S. Qi
State Key Laboratory of Biogeology and Environmental
Geology, China University of Geosciences,
430074 Wuhan, People’s Republic of China

B. De Vivo
Pegaso University, Piazza Trieste e Trento 48,
80132 Naples, Italy

B. De Vivo
Dip. Ambiente e Territorio, Benecon Scarl, Via S. Maria
di Costantinopoli 104, 80138 Naples, Italy

PAHs sources in the study area were mainly of pyrogenic origin, i.e. mostly related to biomass combustion and vehicular emission. On the other hand, values in Sicilian soils seemed to indicate a petrogenic origin, possibly linked to emissions from crude oil combustion and refineries present in the region. Finally, results allowed to calculate the toxicity equivalent quantity (TEQ_{BAP}) levels for the various locations sampled, highlighting that the highest values were found in the Campania region, with 661 and 54.20 ng g⁻¹, in rural and urban areas, respectively. These findings, which could be linked to the presence of a large solid waste incinerator plant, but also to well-documented illegal waste disposal and burning, suggest that exposure to PAH may be posing an increased risk to human health in some of the studied areas.

Keywords Southern Italy · PAHs · Ratio diagnostic · Soil pollution sources · Toxicity equivalent quantity (TEQ_{BAP})

Introduction

Polycyclic aromatic hydrocarbons (PAHs) are diffuse persistent organic pollutants (POPs) that can be found in different environmental media, including air, water and soil. They are human carcinogens, mutagens and are toxic to all living organisms, making them a group of compounds of public concern, which are becoming increasingly studied and monitored in many areas of the world (IARC 1983; Hwang et al. 2003; Nadal et al. 2004; Vane et al. 2014). PAHs are primarily formed through the incomplete combustion of carbon containing fuels such as wood, coal, diesel, fat and tobacco, and most sources of PAHs are anthropogenic, arising from industrial emissions, solid waste incineration and vehicular emissions among others (Dong and Lee 2009). Sixteen US Environmental Protection Agency (EPA) priority PAHs are classified in two main groups of compounds related to the number of aromatic rings: low molecular weights PAHs (LMWPAHs) with 2–3 aromatic rings (naphthalene, acenaphthylene, acenaphthene, fluorene, phenanthrene and anthracene) and high molecular weight PAHs (HMWPAHs) with 4–6 aromatic rings such as fluoranthene, pyrene, benzo[a]anthracene, chrysene,

benzo[b]fluoranthene, benzo[k]fluoranthene, benzo[a]-pyrene, indeno[1,2,3-cd]pyrene, dibenzo[a,h]anthracene, and benzo[g,h,i]perylene. PAHs from a petrogenic source are formed predominantly with those of low molecular weights, whilst the PAHs from a pyrogenic source generally have high molecular weights (Soclo et al. 2000).

Once formed by the mechanisms of partial combustion, PAHs can be found in different media. In particular, soil is considered an important media to quantify PAHs patterns due to its physicochemical properties that allow PAH compounds to be held in soil matrices (Means et al. 1980). PAHs are slightly or completely insoluble in water, and they are adsorbed on soil particles, particularly on soil organic matter. Hence, the physical–chemical properties of soils are responsible for the retention of PAHs in soil matrices. The organic carbon content, the hydrophobicity of soil organic matter and soil texture were estimated to be the most significant parameters controlling the environmental availability of PAHs (Albanese et al. 2015a). Furthermore, some studies (Menzie et al. 1992; Nadal et al. 2004) have demonstrated that the amount of human exposure to PAHs through soils was higher than through air or water. As they exist in different forms with a different degree of toxicity, it is important to characterise individual PAHs compounds as much as possible. However, given their number and variety, often their ratios can be a more effective diagnostic tool to identify potential source patterns and quantify the amount of PAH pollution for a specific area (e.g. Pandey et al. 1999; Yunker et al. 2002; Hwang et al. 2003). One of the most widespread computations used is that involving the LMWPAHs/HMWPAHs ratio introduced by Budzinski et al. (1997) who fixed a value ≤ 1 for pyrogenic source and above 1 for a petrogenic PAHs fingerprint. Since the introduction of this method, other authors have developed alternative takes by using individual PAH compounds, in particular low and high molecular weight groups, to highlight their main source patterns (Hwang et al. 2003, Yunker et al. 2002; Tobiszewski and Namieśnik 2012). Recent development of analytical techniques has seen an increase in the number of studies focusing on individual PAHs compounds, and several studies (Zhang et al. 2006; Albanese et al. 2015a; Islam et al. 2017) have focused on topsoil PAHs occurrence and concentration, helping to shape and informing government policy for human and

ecological safety. For example, Italian environmental law (D. Lgs. 152/2006) establishes threshold values that regulate the mitigation of the PAHs in soil media. This legislation fixes different PAHs concentration values based on the type of PAHs and the land use (e.g. residential and industrial areas). Regulations can guide efforts and resources for reclamation and more detailed monitoring, contextualising interpretations in line with risk-based approaches. However, much needs to be done to establish baselines and understand the mechanisms of these contaminants' movement and availability in the environment. In particular, in southern Italy there exist several potential anthropogenic sources of PAHs such as petroleum exploitation districts, biomass combustion plants, vehicular emissions and residential wood combustion which can all constitute a source of PAHs compounds and contribute to their concentrations in soils and other media. These have been only recently studied, and mainly at a local, small scale, whilst a regional baseline approach has not been carried out yet.

The present study will focus on the 16 (EPA) PAHs priority compounds to carry out a regional survey in southern Italy. PAHs will be characterised in soils of several urban and rural locations to assess their spatial distribution, their potential sources and pathways their level of toxicity.

This study is important because it will constitute the first regional survey carried out in Italy and can be considered a first stepping stone towards a more detailed and meaningful investigation on potential sources and levels of PAHs in the region. It is anticipated that this study will contribute to build a baseline for PAHs characterisations in urban and rural areas of southern Italy. Follow-up studies should be expected in areas where high PAHs concentration levels (i.e. contamination) were found, with a larger number and higher density of (soil and air) samples in each affected location.

Materials and methods

Study areas

The survey area included five administrative regions (Latium, Campania, Basilicata, Calabria and Sicily) of the south of Italy. The total area extends to approximately 81,054 km² with 19.38 million inhabitants,

mostly grouped in urban areas. The overall area is characterised by a principal mountain range, the Apennine chain, which presents very specific geological and morphological features (Fig. 1a).

The main geological features of the study area are a result of the tectonic and orogenic activities related to the nearby boundary between the Eurasian and the African Plates. The area is characterised by active volcanism of different origin, such as that related to the Vesuvius, the Phlegraean Fields, Ischia island, Pontine Islands and Roccamonfina in the Campania region, and the Etna and Aeolian islands in the Northeast of Sicily. Such volcanism can be considered a potential (though limited) source of PAHs during volcanic eruptions (Kozak et al. 2017). From the land use point of view, the study region is devoted to intensive agriculture activities such vineyards, olive plantations—mostly in hilly areas—citrus fruits, seasonal crops and greenhouse products (tomatoes, potatoes, aubergines, peppers and peas) which represent major resources for the local economy (Albanese et al. 2007). In addition to this, large industrial works are also present in the region, processing raw materials of various origins (e.g. petroleum plants, biomass and alloy). Some of these industrial activities can be linked to 'petrogenic' sources of organic pollutants. Some good examples can be found in: (1) the Priolo and Gela districts (Sicily), where the largest petroleum field and refinery in Europe can be found (DPCM 1990); (2) Alta Val d'Agri (Basilicata), which is the location of the largest Italian inland oil field; (3) Crotona (Calabria) and Milazzo (Sicily), the sites of two very large petroleum refineries. An industrial activity that can be linked to potential 'pyrogenic' sources of organic pollutants is that related to energy (and energy-from-waste) generation, which is widely diffused in the study area. Some good examples can be found in: (1) Latium, with the thermal and coal power plants in Civitavecchia and the solid municipal waste incinerator plant in Frosinone; (2) in Campania, with various large industrialised areas within the metropolitan area of Naples, and with the urban solid waste Incinerator plant of Acerra; (3) in Basilicata (Potenza and Melfi) and Calabria (Gioia Tauro), where other incinerators plants can be found (Fig. 2).

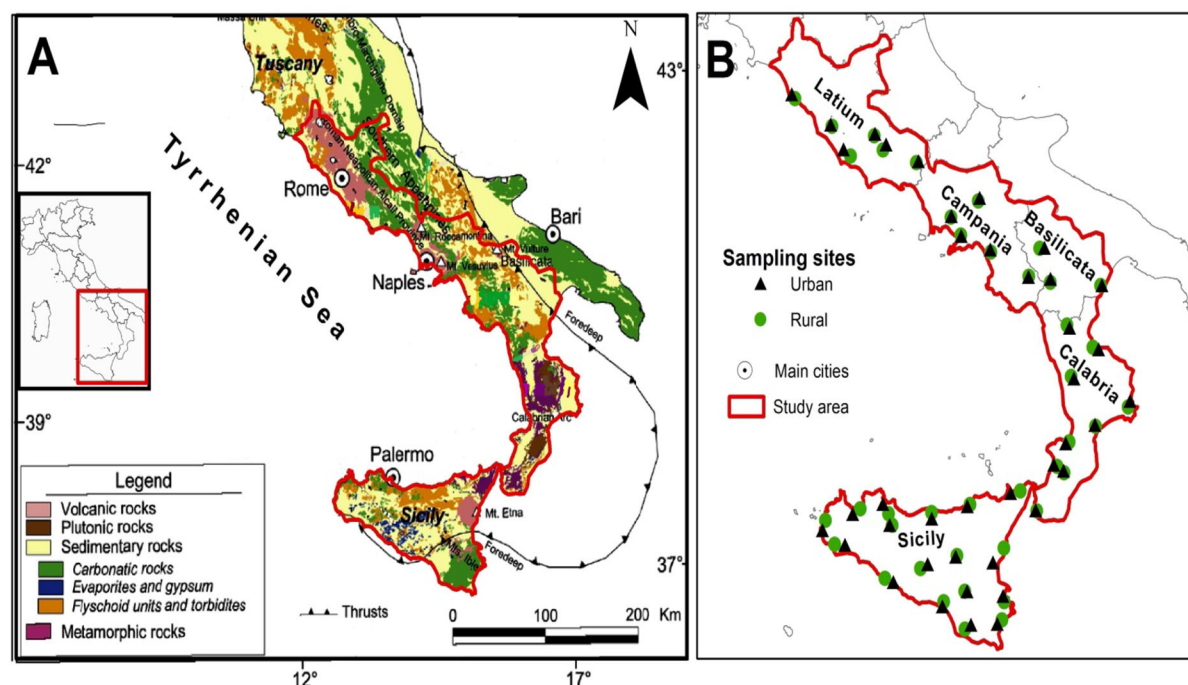


Fig. 1 **a** Simplified Geology of the study area, modified after Doglioni and Flores (1997). **b** Survey area with samples sites

Description of sampling locations

The sampling campaign took place from early April to end of September 2016, with the aim of taking the most representative soil samples in urban and rural areas throughout five administrative regions (Latium, Campania, Basilicata, Calabria and Sicily) of southern Italy. In each region, the main urban areas and the nearest rural areas, where most of the land is devoted to agricultural activities, were selected. The sampling site characterisation and selection was performed using geographical information systems (GIS) data of the industrial and agricultural activities and land use of the study area (ISPRA 2014; ISTAT 2016) as well as using satellite images (Google Earth® professional, 2016 version).

Sample collection and materials

A total of 80 soil samples were collected for polycyclic aromatic hydrocarbons (PAHs) with a nominal density of 2 samples (in urban and rural areas) in each 2500 km² (Fig. 1b). The sampling procedure followed the Geochemical Mapping of Agricultural and Grazing Land Soils (GEMAS) sampling procedure

described by Reimann et al. (2005). All the samples were collected using a stainless steel scoop and were kept in labelled glass bottles and directly stored in ice boxes to minimise the losses caused by volatilisation and initial degradation of the organic compounds (Albanese et al. 2015a). Each topsoil sample (from 0 to 20 cm) was made by homogenising five subsamples at the corners and the centre of a 100 m² square, collecting approximately 1.5 kg in total. The homogenised soil samples were sieved using a < 2 mm mesh sieve after removing stones, detritus and residual roots. Finally, composite samples were stored at −4 °C in the environmental geochemistry laboratory of the University of Naples Federico II (Italy) until instrumental analysis. For each sampling site pH, moisture content and electric conductivity of the soil were measured. Records of the land use, main industrial works and any other human activity observed in the proximity of the sampling locations were recorded at each site, where geographical coordinates were uploaded by global positioning systems (GPS). These attributes would form the dataset subsequently used for the spatial analysis and representation by GIS.

Fig. 2 Land use and main industrial activities in the study area. This industrial repertory is based on the ISPRA (2014) annual report



Sample preparation and analytical procedures

Samples were analysed for PAHs content at the Key Laboratory of Biogeology and Environmental Geology of Ministry of Education, China University of Geosciences, Wuhan, China. For this study, the target analytes were the 16 US EPA priority PAH compounds: naphthalene(Nap), acenaphthylene (Acy),

acenaphthene (Ace), fluorene(Flu), phenanthrene (Phe), anthracene (Ant), fluoranthene (Fluo), pyrene (Pyr), benzo[a]anthracene (BaA), chrysene (Chr), benzo[b]fluoranthene (BbF), benzo[k]fluoranthene (BkF), benzo[a]-pyrene (BaP), indeno[1,2,3-cd]pyrene (IcdP), dibenzo[a,h]anthracene(DahA), and benzo[g,h,i]perylene (BghiP). Ten grams of soil sample was weighed and injected with PAH surrogates

(naphthalene-d8, acenaphthene-d10, phenanthrene-d10, chrysene-d12 and perylene-d12) and Soxhlet-extracted with dichloromethane (DCM) for 24 h. The extracts were treated with activated copper granules to remove elemental sulphur, concentrated and solvent-exchanged to n-hexane and further reduced to 2–3 ml by a rotary evaporator (Heidolph 4000, Germany). A 1:2 (v/v) alumina/silica gel column (450 °C muffle drying for 4 h, both 3% deactivated with H₂O before using) was used to clean up the extracts, and PAHs were eluted with 70 ml of DCM/hexane (2:3). The eluate was then reduced to 0.2 ml under a gentle stream of nitrogen. A known quantity of hexamethylbenzene was added as an internal standard for PAHs analysis prior to instrumental quantitation for the PAHs.

PAHs were analysed using GC-MS (Agilent 6890 N/5975 MSD) coupled with a HP-5972 mass selective detector operated in the electron impact mode (70 eV) installed with a DB-5 capillary column (30 m × 0.25 mm diameter, 0.25 µm film thickness). Helium (99.999%) was used as the GC carrier gas at a constant flow of 1.5 ml min⁻¹. An 1 µl concentrated sample was injected with splitless mode. The chromatographic conditions were as follows: injector temperature 270 °C; detector temperature 280 °C; oven temperature initially at 60 °C for 5 min, increased to 290 °C at 3 °C min⁻¹, and held for 40 min. Chromatographic peaks of samples were identified by mass spectra and retention time.

Quality control

Procedure types used for quality assurance and quality control (QA/QC) were as follows: method blank control (procedural blank samples), parallel sample control (duplicate samples), solvent blank control and basic matter control (US EPA 2002). In order to ensure the validity of the analyses during the experiment, different reagents and procedures were used:

1. One thousand nanograms (ng) of naphthalene-D8, acenaphthene-D10, phenanthrene-D10, chrysene-D12 and perylene-D12 were used as recovery surrogates, and 1000 ng hexamethylbenzene was added in extracts as the internal standard substance. The spiked recoveries of PAHs using composite standards were $79.9 \pm 14.5\%$ for naphthalene-D8, $74.2 \pm 9.4\%$ for acenaphthene-

D10, $91.5 \pm 11.6\%$ for phenanthrene-D10, $87.1 \pm 8.5\%$ for chrysene-D12 and $89.2 \pm 11.0\%$ for perylene-D12, respectively.

2. An internal standard method was used for quantification: a six-point calibration curve was established according to the results from the PAH-16 standard reagents with concentration of 10, 5, 2, 1, 0.5 and 0.2 mg l⁻¹. For PAHs, the target compounds were identified on the basis of the retention times and selected quantitative ion.
3. During the pre-treatment, a procedural blank and a parallel sample consisting of all reagents were run to check for interference and cross-contamination in every set of samples (about 16 samples). Only low concentrations of few target compounds can be detected in procedural blank samples. For more than 96% of target compounds in parallel samples, the relative error (RE, %) of concentrations is less than 50%, which is acceptable for Specification of Multi-purpose Regional Geochemical Survey and Guidelines for sample analysis of Multi-purpose Regional Geochemical Survey recommended by China Geological Survey (DD2005-1 and DD2005-3).
4. During the GC-MS analysis period, a solvent blank sample and a PAH-16 standard reagent with concentration of 5 mg l⁻¹ and 100 µg l⁻¹ were injected every day before analysing the soil samples. The target compounds were not detectable in the solvent blank samples.
5. Multi-injections were used for precision or accuracy. The samples of different concentrations were injected continually for ten times, and the relative standard deviation was calculated. RSD for all the target compounds ranged from 3.2 to 7.9%. The final concentrations of PAHs in all samples were corrected according to the recovery of the surrogates, and the results of blank samples were subtracted.

Statistical computations

Univariate and multivariate analyses were carried out on the 16 PAHs through descriptive statistics and compositional principal factor analysis modelling, which helped displaying the variation of these compounds and their main correlations in the survey area. Computations and graphical representations were

implemented by mean of the open source statistical software R (Templ et al. 2011).

Source apportionment

PAHs diagnostic ratios

Whilst several PAHs diagnostic ratios are available in the literature, (e.g. Katsoyiannis et al. 2007; Ravindra et al. 2008), in this study, four specific PAH molecular ratios were used for the identification of PAHs pollution sources: LMWPAHs/HMWPAHs, Fluo/(Fluo + Pyr), IcdP/(IcdP + BghiP) and BaA/(BaA + Chr) (Figs. 5, 6; Table 1). In particular, the reasons why these were chosen are that: the ratio LMWPAHs/HMWPAHs ≤ 1 corresponds to pyrogenic sources, and > 1 for petrogenic sources (Budzinski et al. 1997); the ratio Fluo/(Fluo + Pyr) < 0.4 has been shown to indicate petroleum source, between 0.4 and 0.5 implies fossil fuel combustion, and a ratio > 0.5 is the characteristic of biomass and coal combustion (Yunker et al. 2002). IcdP/(IcdP + BghiP) < 0.2 is an indication of petroleum sources, whilst that between 0.2 and 0.5 indicates that the PAHs usually derive from petroleum combustion (liquid fossil fuel, vehicle and crude oil combustion) and IcdP/(IcdP + BghiP) > 0.5 strongly indicates the contribution of coal, grass and wood combustion (Tobiszewski and Namieśnik 2012). Yunker et al. (2002) implemented the BaA/(BaA + Chr) ratio and the value < 0.2 to mark a petroleum/petrogenic source, whilst that

between 0.20 and 0.35 is linked to a combustion and > 0.35 is related to a traffic emission.

Compositional multivariate computation: factor analysis

R-mode factor analysis, a type of multivariate statistics, was chosen to explain the correlation structure of the 16 EPA PAHs compounds (variables) using a smaller number of factors (Reimann et al. 2002). This methodology has been shown to successfully correlate the PAHs distribution to their main hypothetical origins (Albanese et al. 2015a; Islam et al. 2017). To minimise and/or eliminate the presence of outliers and spurious correlation (Pawlowsky-Glahn and Buccianti 2011), isometric log-transformed data (ilr) are recommended in this type of multivariate analysis (Filzmoser et al. 2009). In order to facilitate the interpretation of results, varimax rotation was used, since it is an orthogonal rotation that minimises the number of variables that have high loadings on each factor, simplifying the transformed data matrix and assisting interpretation (Reimann et al. 2002). The different factors obtained were studied and interpreted in accordance with their presumed origin (petrogenic, pyrogenic or mixed) (Table 4, Fig. 7). The factor score values were mapped at each sample site using GIS software GeoDAS (Cheng et al. 2001) and ArcGIS (ESRI 2012). GeoDASTM was used to produce interpolated geochemical maps of the factor scores by means of the multifractal inverse distance weighted (MIDW) algorithm (Cheng et al. 1994; Lima

Table 1 Compilation of PAH ratios and molecular markers for source diagnosis

Compounds, compounds ratios	Values, ranges	Sources	References
LMWPAHs/HMWPAHs	≤ 1	Pyrogenic combustion	Budzinski et al. (1997)
	≥ 1	Petrogenic source	Budzinski et al. (1997)
	< 0.40	Petroleum/petrogenic source	Yunker et al. (2002)
Fluo/(Fluo + Pyr)	0.4–0.5	Fossil fuel combustion	Yunker et al. (2002)
	> 0.5	Biomass and coal combustion	Yunker et al. (2002)
	< 0.20	Petroleum/petrogenic source	Yunker et al. (2002)
BaA/(BaA + Chr)	0.2–0.35	Coal combustion	Yunker et al. (2002)
	> 0.35	Traffic emission	Yunker et al. (2002)
	< 0.20	Petrogenic source	Tobiszewski and Namieśnik (2012)
IcdP/(IcdP + BghiP)	0.20–0.50	Fuel combustion	Tobiszewski and Namieśnik (2012)
	> 0.50	Coal and biomass combustion	Tobiszewski and Namieśnik (2012)

et al. 2003). MIDW is one of the most widely used interpolation methods on geochemical data because it preserves high-frequency information, retains local variability taking into consideration both spatial association and local singularity (Cheng et al. 1994, 2001; Lima et al. 2003). Singularity is an index representing the scaling dependency from multifractal point of view, which characterises how the statistical behaviour of a spatial variable changes as the measuring scale changes (Cheng et al. 1994). Spatial association represents a type of statistical dependency of values at separate locations, and its indexes (e.g. covariance, autocorrelation and semivariogram) have been used to characterise the local structure and variability of surfaces (Cheng et al. 1994). During interpolation and mapping of geochemical variables, both spatial association and scaling are taken into account. Despite the low density, interpolation is still a valid tool at regional level, as shown, for example, by the production of European geochemical atlases, which have used similar techniques (Reimann et al. 2012; Ottesen et al. 2013; Albanese et al. 2015b). The concentration–area (C–A) fractal method (Cheng et al. 1994) was applied to set the factor score intervals (Thiombane et al. 2017) in interpolated surface images generated by the MIDW method, and ArcGISTM software was used for the graphical presentation of the results (Figs. 8, 9, 10).

Characterisation and toxicity assessment of PAHs

For each PAH compound, the toxicity equivalency factor (TEF) was established to allow measuring its relative carcinogenicity (EPA 1984; Nisbet and Lagoy 1992). Among the 16 EPA PAHs, seven of them, including BaA, Chr, BaP, BbF, BkF, IcdP and DahA, present high toxic and carcinogenic effects. In particular, benzo[a]-pyrene (BaP) is considered as one of the most toxic PAHs and it has been used to quantify the relative toxicity of others PAHs (Nisbet and Lagoy 1992), also because it is the only compound with sufficient toxicological data to derive carcinogenic factors among all other potentially carcinogenic PAHs. The toxicity of soil can be measured using the BaP toxic equivalent quantity (TEQ_{BaP}) (Nadal et al. 2004) for each sampling site using the equation described below:

$$\text{TEQ}_{\text{BAP}} = \sum_{i=1}^7 \text{TEF} \times C_{\text{PAH}i} \quad (1)$$

where TEQ_{BAP} is the toxic equivalent quantity of *i*th PAH, the TEF of BaA, Chr, BaP, BbF, BkF, IcdP and DahA corresponds to 0.1, 0.01, 1, 0.1, 0.1, 0.1 and 1, respectively, and $C_{\text{PAH}i}$ is the concentration of the *i*th PAH in the soil.

Results and discussions

Variety PAHs concentrations in the survey area

Table 2 shows the descriptive statistic of the 16 EPA PAHs compounds in soils found on the sample locations of the studied area.

The total concentration of the 16 PAHs in urban and rural area ranged from 7.62 to 755 ng g^{−1} with a mean value of 84.85 ng g^{−1}, and from 1.87 to 11,353 ng g^{−1} with a mean of 333 ng g^{−1}, respectively. The spatial distribution of the \sum 16 PAHs concentrations in the survey areas can be seen in Fig. 3, which compares urban and rural areas using proportional thematic mapping.

Figure 3a shows high total PAHs values (ranging from 59.52 to 755 ng g^{−1}) in correspondence with urban towns such as Rome (755 ng g^{−1}), Naples (715 ng g^{−1}) and Palermo (303 ng g^{−1}). These areas are also the most densely populated and urbanised cities of the southern Italy. In a recent study, Albanese et al. (2015a) highlighted that the principal source of the PAHs pollutants in the Neapolitan (Campania) soils is related to pyrogenic combustion (vehicular emission). When considering the variation of total PAHs in rural areas (Fig. 3b), the highest values ranged from 33.30 to 11,353 ng g^{−1} and are found mostly in the sampling sites of the Campania (Sarno Basin with 11,353 ng g^{−1}, Acerra with 917 ng g^{−1} and Battipaglia with 276 ng g^{−1}), and the Sicily countryside (Acireale with 74.25 ng g^{−1}, Trapani with 56.56 ng g^{−1} and Milazzo with 52.45 ng g^{−1}). When compared to other studies of PAHs in urban and rural soils worldwide, the ranges reported in our study present similar concentrations ranges, though with some noticeable differences (Table 3). Morillo et al. (2007) found total PAHs values ranging from 148 to 3410 ng g^{−1} in the Turin (Italy) urban area; these

Table 2 Descriptive statistic of the 16 US EPA PAHs compounds from the survey area

PAHS (ng g ⁻¹)	Abbreviation	DL ^a	TEF ^b	Urban				Rural			
				Min	Mean	Median	Max	Min	Mean	Median	Max
Naphthalene	Nap	0.013	0.001	0.39	2.49	2.01	9.83	0.33	3.21	2.05	25.13
Acenaphthylene	Acy	0.028	0.001	0.08	0.60	0.20	8.07	0.03	1.41	0.21	22.78
Acenaphthene	Ace	0.047	0.001	0.04	0.24	0.11	2.40	0.03	0.79	0.11	25.20
Fluorene	Flo	0.009	0.001	0.12	0.61	0.57	1.49	0.11	1.21	0.53	25.05
Phenanthrene	Phe	0.001	0.001	0.85	6.74	3.34	47.32	0.40	17.87	2.80	570
Anthracene	Ant	0.005	0.01	0.04	0.77	0.19	8.57	0.01	3.78	0.14	131
Fluoranthene	Fluo	0.003	0.001	0.44	10.43	1.68	108	0.13	25.22	1.57	829
Pyrene	Pyr	0.003	0.001	0.40	10.74	1.60	113	0.11	30.98	1.61	1059
Benz[a]anthracene	BaA	0.002	0.1	0.25	7.27	1.04	93.30	0.19	26.86	0.98	950
Chrysene	Chr	0.004	0.01	0.41	8.58	1.84	93.15	0.10	33.04	1.85	1101
Benzo[b]fluoranthene	BbF	0.004	0.1	0.22	15.05	2.75	151	0.07	86.58	2.43	3055
Benzo[k]fluoranthene	BkF	0.004	0.1	0.19	5.09	0.99	51.38	0.05	19.38	0.84	633
Benzo[a]pyrene	BaP	0.001	1	0.12	10.47	1.32	112	0.08	70.14	1.14	2510
Indeno[1,2,3-c,d]pyrene	IcdP	0.003	0.1	0.04	2.64	0.42	33.51	0.01	7.91	0.31	226
Dibenzo[a,h]anthracene	DahA	0.003	1	0.01	0.50	0.15	6.09	0.01	1.45	0.09	41.23
Benzo[g,h,i]perylene	BghiP	0.001	0.01	0.11	2.65	0.71	33.38	0.03	6.28	0.44	161
∑16 PAHs				7.62	84.85	18.65	755	1.87	336	18.81	11,353

Minimum (Min), mean, median, maximum (Max) concentration values are expressed in ng g⁻¹

^aDetection limit of the 16 PAHs compounds

^bToxicity equivalency factor (TEF) (Nisbet and Lagoy 1992)

values are four times greater than those found in urban areas of this survey study. The total 16 PAHs found in others European urban cities such in Seville (ranging from 89.5 to 4004 ng g⁻¹; Morillo et al. 2008), London (ranging from 400 to 67,000 ng g⁻¹; Vane et al. 2014), Glasgow (ranging from 48 to 51,822 ng g⁻¹; Morillo et al. 2007) and Moscow (208–3880 ng g⁻¹; Agapkina et al. 2007) are all higher than the values displayed in the urban areas of this study. Similarly, though at a larger scale, the ranges shown in the urban areas of Beijing (China) (Tang et al. 2005) and Delhi (India) (Bhupander et al. 2012) went from 219 to 27,825 ng g⁻¹ and from 81.6 to 45,017 ng g⁻¹, respectively. On the other hand, the rural areas of the southern Italy showed higher total concentration of PAHs compared to the values displayed in Hong Kong (China) (Zhang et al. 2006) and Delhi (Agarwal et al. 2009), ranging from 42.3 to 410 ng g⁻¹, and from 830 to 3880 ng g⁻¹, respectively. Many studies revealed that PAHs pollution sources are usually related to fuel combustion from

traffic vehicle which is mostly occurring in urban areas. These studies are confirmed by the findings on some urban areas of the present study (e.g. high 16 PAHs in urban areas of Latium, Naples and Palermo). However, this survey also highlighted some unexpected higher values outside of the urban areas, perhaps due to the presence of industrial activities such as incinerator plants, oils refineries as well as and illegal activities (e.g. wood and solid waste burning in the Acerra district and in the Naples wider metropolitan area). Even though such large industrial activities are actually forbidden in urban areas by the most recent Italian environmental legislation, their influence on the distribution of PAHs on rural areas remained, until now, largely unexplored.

These findings are confirmed by the Tukey's box-and-whiskers plots (Fig. 4), which display how all 16 PAHs compounds showed higher concentration values in the rural sampling sites compared to the urban ones. On the other hand, Flo, Phe, Ant, Fluo, BaA, BbF, BkF, BaP, IcdP, DahA and BghiP, which are all

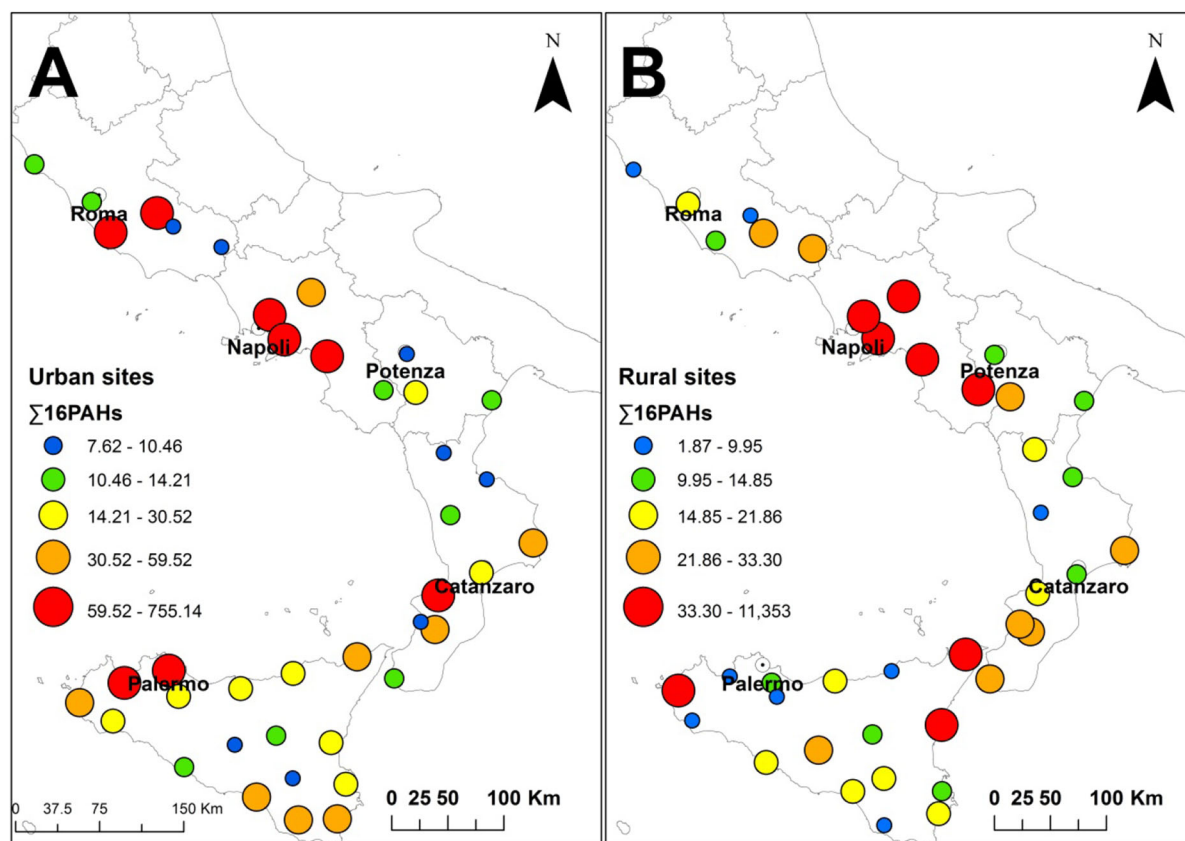


Fig. 3 Dots maps of the 16 PAHs concentration in urban (a) and rural (b) soils

Table 3 Total PAH concentrations (ng g^{-1} dry weight) in the survey area topsoil compared to those found in other studies in the recent literature

Study areas	Type of study area	Σ PAHs in soils (ng g^{-1})	Authors
Southern Italy	Urban areas	7.62–755	This study
Southern Italy	Rural areas	1.87–11,353	This study
Turin, Italy	Urban areas	148–3410	Morillo et al. (2007)
Hong Kong, China	Rural areas	42.3–410	Zhang et al. (2006)
Delhi, India	Urban areas	81.6–45,017	Bhupander et al. (2012)
Delhi, India	Rural areas	830–3880	Agarwal et al. (2009)
Beijing, China	Urban areas	219–27,825	Tang et al. (2005)
Moscow, Russia	Urban areas	208–9604	Agapkina et al. (2007)
Glasgow, UK	Urban areas	48–51,822	Morillo et al. (2007)
London, UK	Urban areas	400–67,000	Vane et al. (2014)
Seville, Spain	Urban areas	89.5–4004	Morillo et al. (2008)

HMWPAHs, displayed higher median concentration values in urban areas (Table 2). Since the median is a good, intuitive metric of centrality representing a ‘typical’ or ‘middle’ value (Reimann et al. 2008), it is reasonable to infer that HMWPAHs are therefore more likely to be related to those pollution sources that are

occurring mostly in urban areas, such as vehicular emission and fuel combustion, whereas LMWPAHs could be more strongly related to biomass and oil combustion sources which are more likely to be occurring in rural areas (Chen et al. 2005; Aichner et al. 2007).

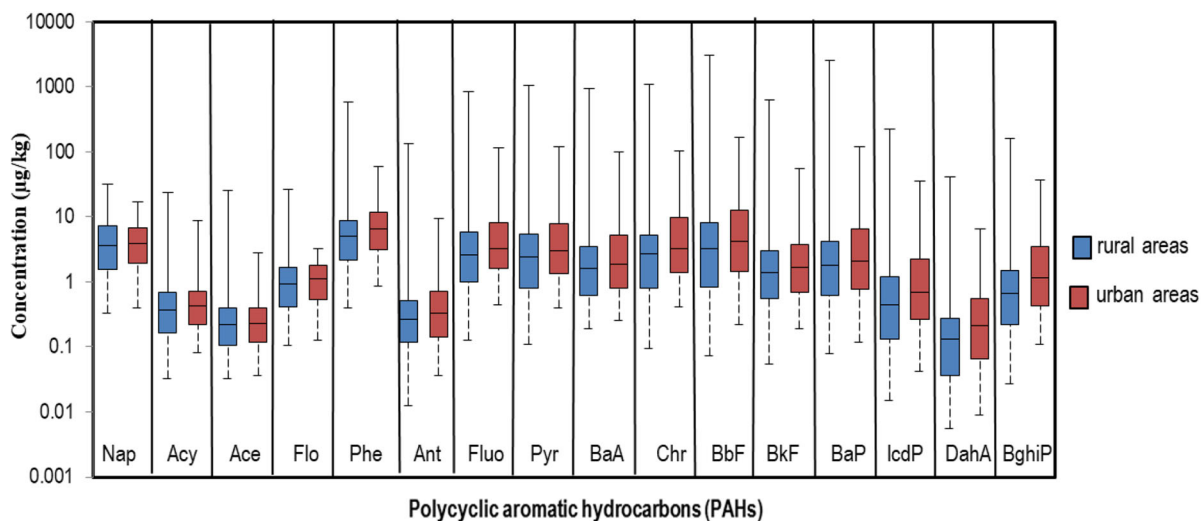


Fig. 4 Tukey's box-and-whiskers plot of individual PAH concentration (ng g^{-1}) in the survey area

Diagnostic ratios and source apportionment of PAHs compounds

The LMWPAHs/HMWPAHs ratios (Fig. 5) were interpreted in terms of source apportionment. Soils from rural Sicilian areas displayed a higher mean value compared to those from urban areas, with ratios ranging from 0.12 to 2.48 (mean value of 1.14) compared to 0.08 to 1.35 (mean value of 0.56). These values also indicate that the most likely sources of PAHs in Sicilian rural areas may be related to 'petrogenic' emissions. As mentioned earlier, in Sicily there are some of the most important Italian oil fields and refineries (e.g. Priolo, Gela, Ragusa and Milazzo—Bevilacqua and Braglia 2002). These industrial activities can give rise to 'petrogenic' emissions of PAHs and are found in suburban or nearby rural areas.

In Calabria, the ratios are ranging from 0.11 to 1.37 (mean value 0.73) and from 0.25 to 1.11 (mean value 0.80) in urban and rural areas, respectively. In comparison, in Basilicata the ratios are ranging from 0.49 to 1.11 (mean value 0.59) and from 0.25 to 1.55 (mean value 0.71) in urban and rural areas, respectively. From these values, it seems that Calabria displays some similar or slightly higher average values of LMWPAHs/HMWPAHs ratios compared to those in Basilicata. A potential explanation is that in Calabria there may be more sources of LMWPAHs (petrogenic, e.g. the large Crotone oil refineries) than

in Basilicata. However, the highest measured ratios in Basilicata were found in the urban (ratio = 0.99) and rural (ratio = 1.55) areas of Viggiano municipality, in proximity of the most important Italian inland petroleum exploitation (Alta Val d'Agri oil field) (ISPRA 2014).

Campania and Latium presented lower ranges of LMWPAHs/HMWPAHs in urban areas, from 0.08 to 1.02 (mean value of 0.29) and from 0.08 to 0.52 (mean value of 0.27), respectively, and, from 0.08 to 0.14 (mean value of 0.10) and from 0.16 to 0.96 (mean value of 0.42) in rural areas, respectively. Given that most of the sites (but one) show a LMWPAHs/HMWPAHs ratio < 1, it seems that in these two areas the most likely sources of PAHs may be related to pyrogenic activities.

Individual molecular compound ratios were used to evaluate their potential sources (Hwang et al. 2003, Yunker et al. 2002). Fluo/(Fluo + Pyr) ratios ranged from 0.39 to 0.64, and 0.40 to 0.61 in urban and rural areas, respectively (Fig. 6a). When considering sources characterisations ranges, it appeared that none of the sampling sites, neither urban nor rural areas, presented ratios characteristic of petrogenic sources (ratio < 0.4). Similarly, very limited sites displayed BaA/(BaA + Chr) ratio below 0.2. On the other hand, about 48.3 and 34.5% of the urban and rural sampling sites, respectively, presented a ratio that could be linked to a fuel combustion source (ratios from 0.4 to 0.5). This is also backed up by looking at two other

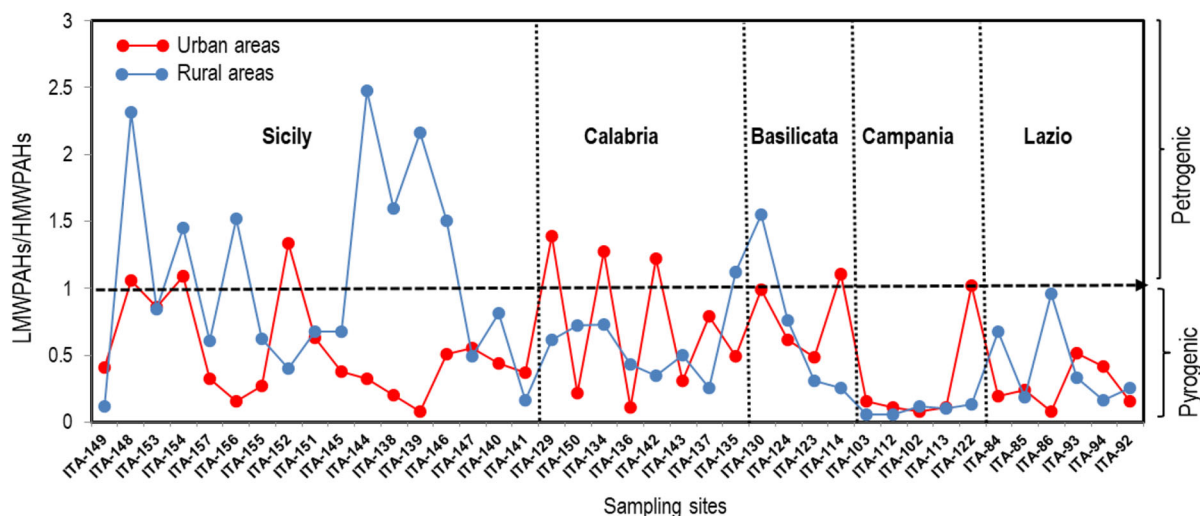


Fig. 5 LMPAHs/HMPAHs ratios scatter diagram of each sample site location. The different symbology (dots) reflects whether the sites were urban or rural nature

diagnostic ratios: (1) the IcdP/(IcdP + BghiP) ratio, for which the results showed that the majority of the urban (58.2%) and rural (52.8%) areas fall in the range (< 0.5) characterised by fuel combustion sources; (2) the BaA/(BaA + Chr) ratio (Fig. 6b), for which results indicated that the majority of the samples sites, in urban and rural areas, presented ratios > 0.35 , corresponding to the same traffic combustion sources. Besides, Fluo/(Fluo + Pyr) ratios highlighted that 51.7 and 65.5% of urban and rural areas are displayed in the plot (ratios above 0.5) where sources of PAHs are more likely to be related to coal and biomass combustion. These measurements seem to suggest that the most likely sources of PAHs compounds in the studied areas may be related to fuel (vehicular and biomass) combustion (pyrogenic).

To summarise, by using different molecular diagnostic ratios, it can be highlighted that:

- LMPAHs/HMPAHs ratios \rightarrow dominant pyrogenic sources ($r < 0.1$)
- LMPAHs/HMPAHs ratios \rightarrow some petrogenic sources ($r > 1$ – Sicily).
- Fluo/(Fluo + Pyr) ratio \rightarrow no petrogenic source ($r < 0.4$)
- BaA/(BaA + Chr) ratio \rightarrow dominant pyrogenic sources ($0.4 < r < 0.5$)

Factor score maps for sources patterns

The total variance expressed by the 16 PAHs variables was 73.05% through three-factor models, F_1 , F_2 and F_3 , accounting for 46.65, 15.05 and 11.35% of the variance, respectively (Table 4). Variables with loadings > 0.50 were considered to describe the main composition of each factor. All variables hold communalities > 0.5 (50% of variability), which means that they were all well correlated to one another. The associations of PAHs compounds for the three-factor models, sorted in descending loading values, were:

F_1 : Fluo, Pyr, BaA, Chr, - (Nap, Acy, Flo)

F_2 : BkF, BbF, BaP, - (Nap, Acy, Ace)

F_3 : BghiP, DahA, IcdP, - (Ant)

The 16 PAHs variables were organised in two groups by using the log-transformed data and a varimax rotation in the factor analysis (Table 4), allowing to distinguish between positive and negative correlations within the three-factor models: G_1 = Fluo, Pyr, BaA, Chr, BbF, BkF, BaP, IcdP, DahA and BghiP (positive correlation) and G_2 = Nap, Acy, Flo, Nap, Acy, Ace and Ant (negative correlation). These compound associations actually matched the main two groups of PAHs, where G_1 corresponds to the high molecular weight PAHs (HMPAHs) and G_2 to the low molecular weight PAHs (LMWPAHs). In addition to this, each of the three-factor models displayed variables associations which can be used

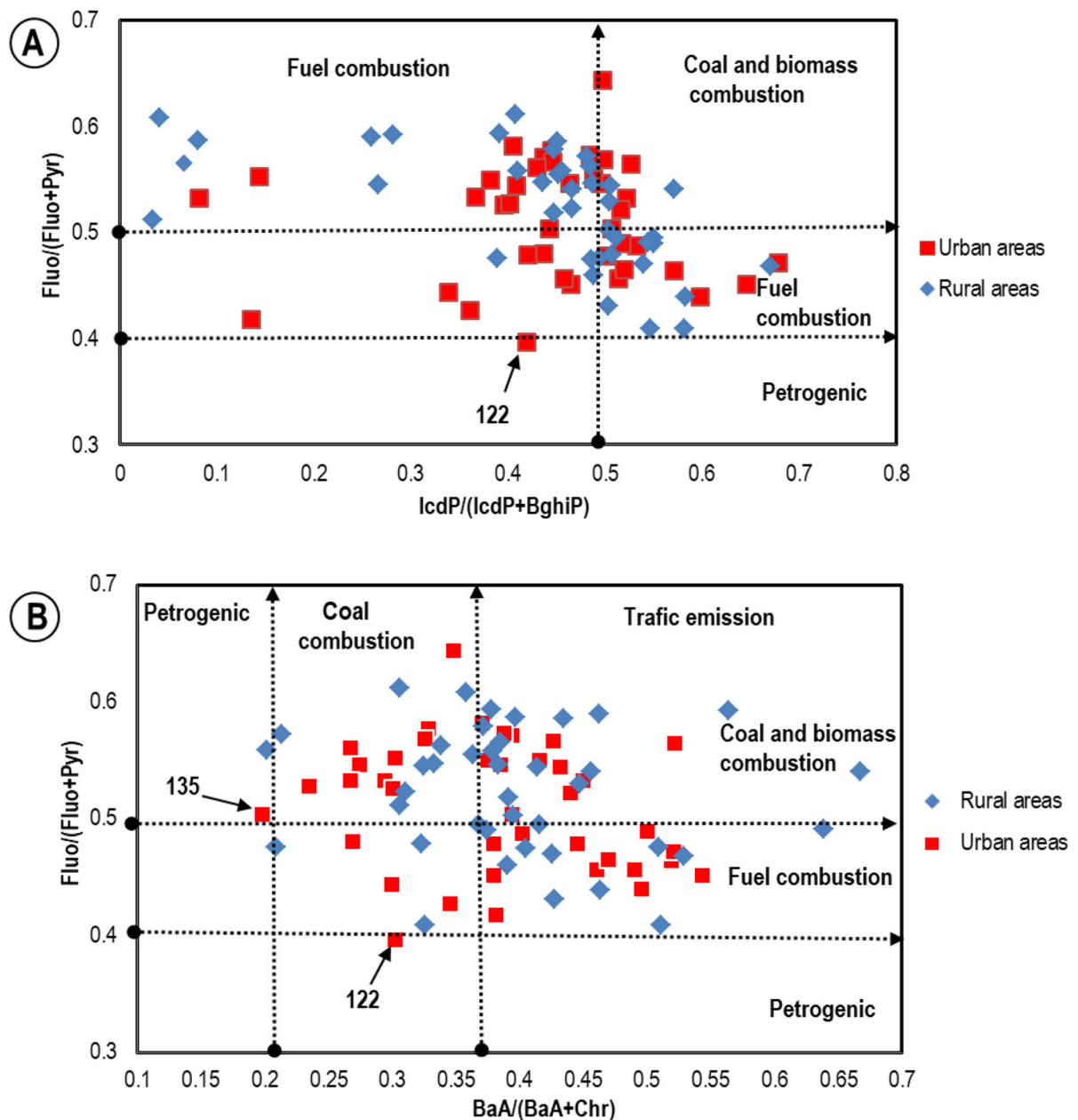


Fig. 6 Cross-plots for the isomeric ratios: **a** IcdP/(IcdP + BghiP) versus Fluo/(Fluo + Pyr) and **b** BaA/(BaA + Chr) ratio against Fluo/(Fluo + Pyr)

to further investigate and reveal the potential sources of the 16 PAHs contaminants in the survey area (Fig. 7).

The F_1 association (Fluo, Pyr, BaA, Chr, - (Nap, Acy, Flo)) accounted for the highest total variance (46.65%) with good adequacy (eigenvalues = 7.46 > 1) between the factor and its variables.

Factor scores distributions were processed in a GIS environment and interpolated to be displayed in maps to better visualise regional PAHs distribution (Lima et al. 2003; Thiombane et al. 2017). The F_1 factor scores interpolated map (Fig. 8) presented the highest values (from 1.13 to 2.77) in and around the largest urban sites (Rome, Naples and Palermo). These areas

Table 4 Varimax-rotated factor (three-factor model) of isometric logratio clr back-transformed variables for 80 topsoil samples from the survey area; bold entries: loading values over 0.50

Variables	Factors			Communalities
	F1	F2	F3	
Nap	− 0.72	− 0.58	− 0.03	0.86
Acy	− 0.51	− 0.13	− 0.42	0.51
Ace	− 0.32	− 0.66	− 0.47	0.75
Flo	− 0.55	− 0.74	− 0.27	0.93
Phe	− 0.29	− 0.73	− 0.27	0.69
Ant	0.10	− 0.07	− 0.74	0.56
Fluo	0.92	− 0.08	0.03	0.86
Pyr	0.92	0.27	− 0.06	0.92
BaA	0.81	0.41	− 0.10	0.83
Chr	0.71	0.30	0.28	0.67
BbF	0.30	0.79	0.21	0.75
BkF	− 0.13	0.82	− 0.05	0.69
BaP	0.44	0.76	− 0.02	0.78
IcdP	0.24	0.38	0.69	0.68
DahA	− 0.19	0.31	0.69	0.61
BghiP	0.25	− 0.22	0.73	0.65
Eigenvalues	7.464	2.408	1.816	
Total variance in %	46.652	15.05	11.348	
Cum. of total variance (%)	46.652	61.703	73.051	

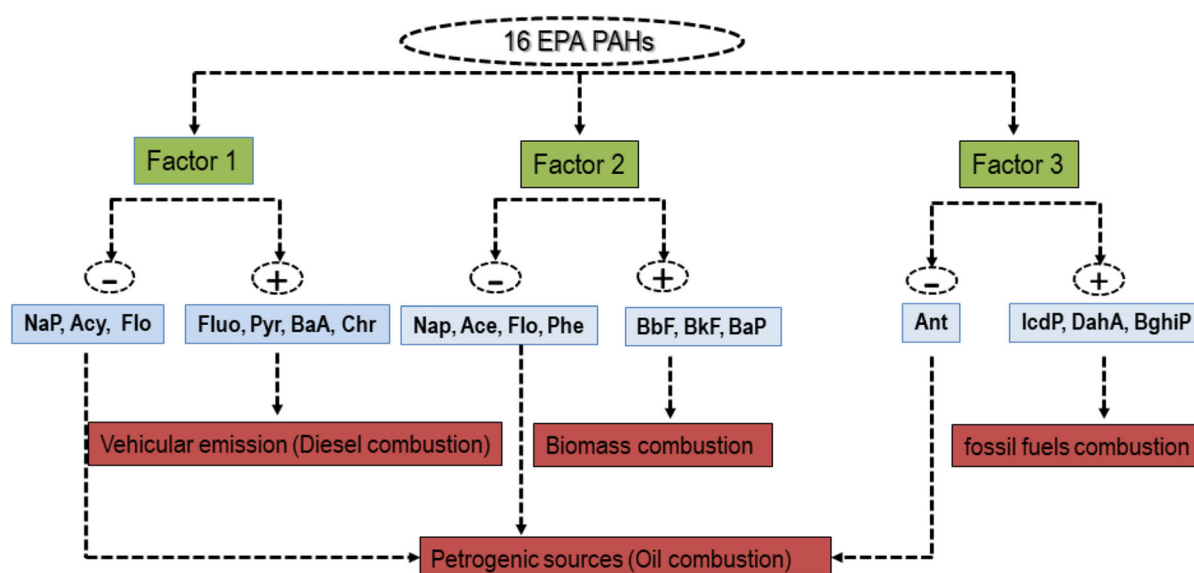


Fig. 7 Flowchart displaying the possible pollution sources of PAHs in the south of Italy throughout factor analysis. Each factor is characterised by antithetic variables association

(marked by − and + symbols) which reveal the possible sources of PAHs in the survey area

are characterised by substantial vehicular traffic and emission (Spaziani et al. 2008). The lowest factor scores loadings (from − 2.52 to − 0.84) are mostly

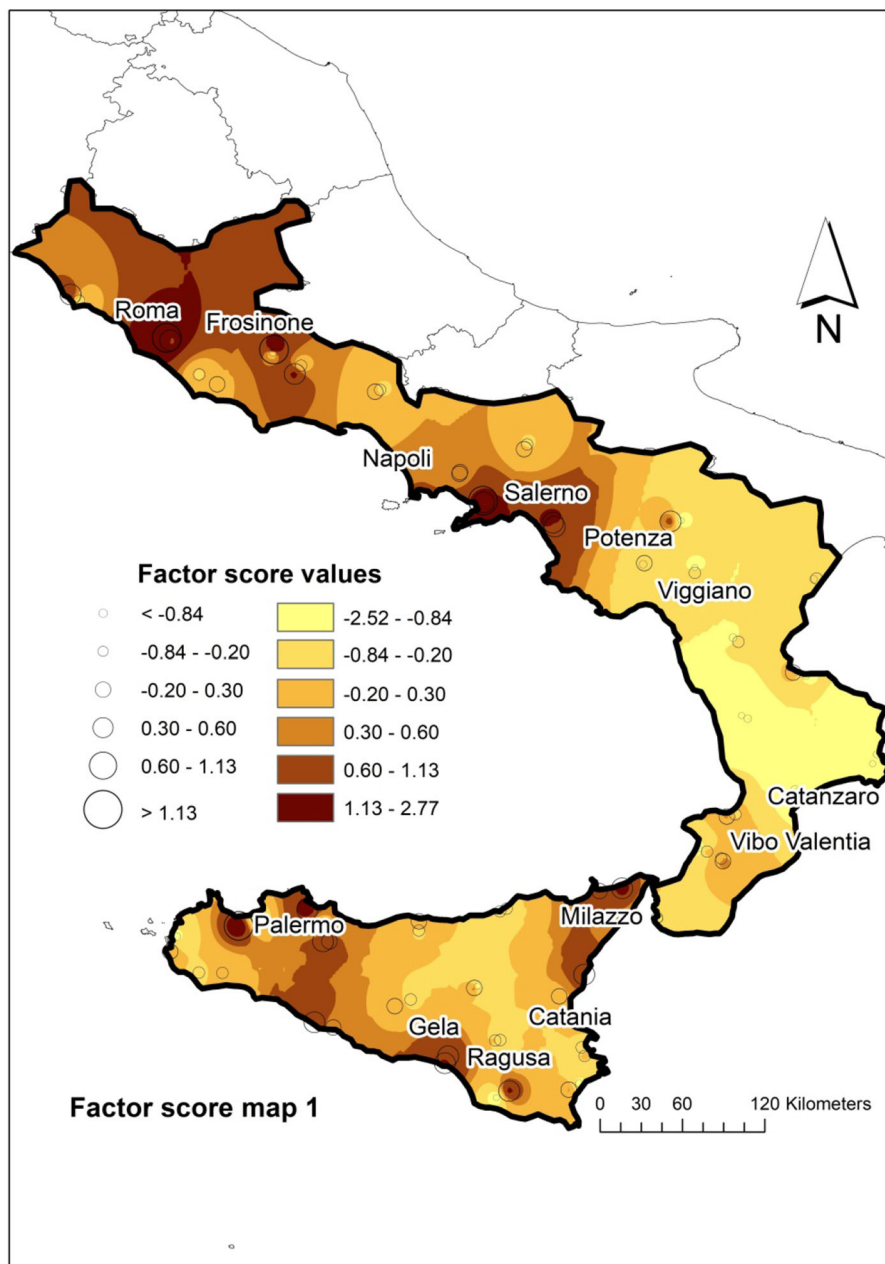
distributed in and around the rural sites in Basilicata (Val d'Agri oil field, Viggiano municipality), Calabria and Sicily (near the Priolo, Gela and Ragusa districts).

Some of the compounds part of this factor (Fluo, Pyr, BaA and Chr) are contaminants usually related to traffic and vehicle exhaust emissions (Sofowote et al. 2008; Li et al. 2012), which are obviously most likely to occur in urban areas. On the other hand, the other compounds part of this factor (Nap, Acy and Flo) are instead usually indicative of spilled-oil-related products at low-temperature combustion (Yunker et al. 2002; Hwang et al. 2003; Bucheli et al. 2004) and can

be associated with similar ‘petrogenic’ sources (e.g. oil combustion from petroleum exploitation industries) in the Basilicata and Sicily regions.

The F_2 association (BbF, BkF, BaP, - (Nap, Acy, Ace)) expressed 15.05% of the total variance with an eigenvalue of 2.41. The F_2 factor scores map (Fig. 9) presented the higher factor scores values (from 1.15 to 1.81) mostly distributed in and around the rural sites within the Campania region. In particular, the highest

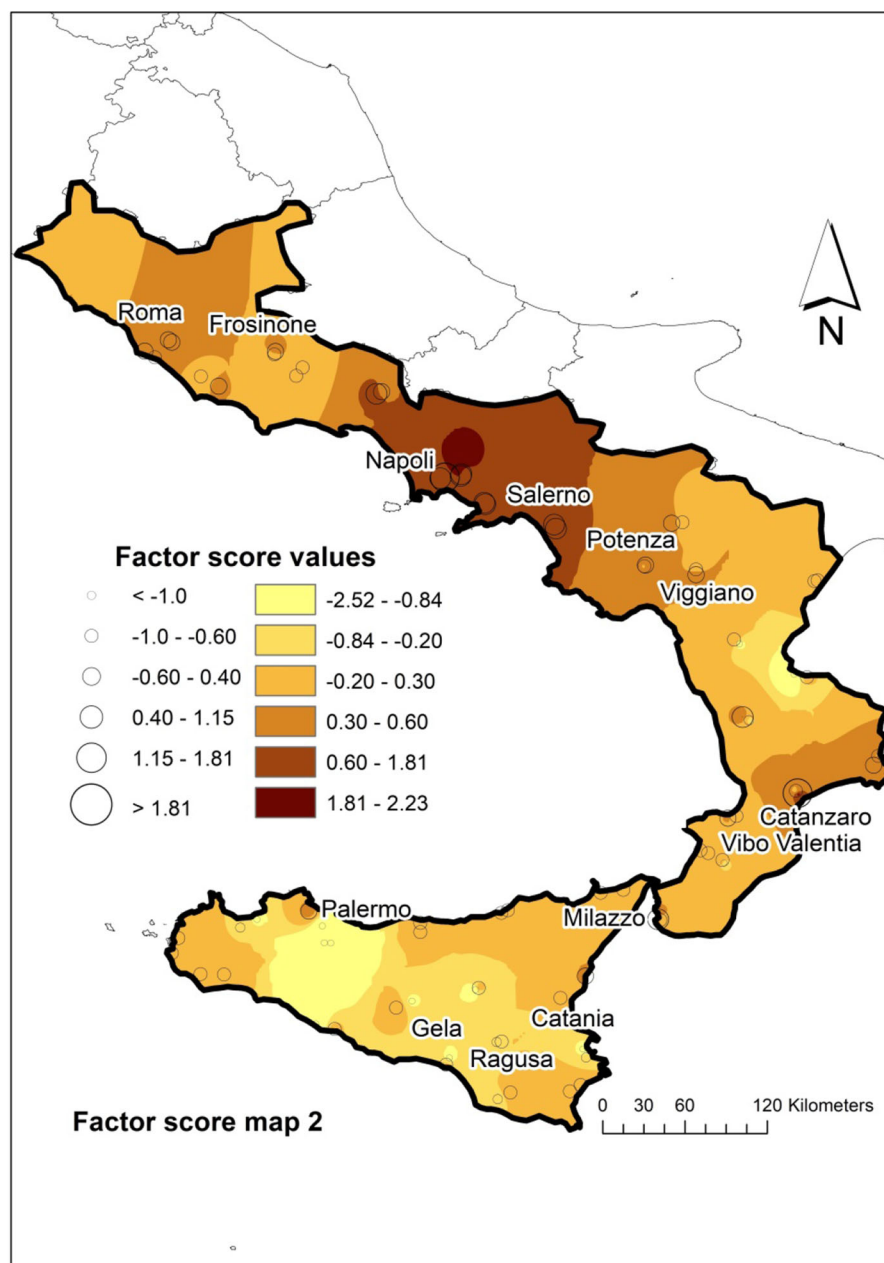
Fig. 8 Interpolated factor score map of the factor 1 (F1). Factor score values ranges are created by means of fractal concentration–area plot (C–A method)



factor score values (from 1.81 to 2.23) were noted in the Acerra district. This municipality, which falls within the metropolitan/suburban area of Naples, is characterised by the presence of a large solid waste incinerator plant, but also by some illegal waste disposal and burning, and illegal practice of industrial toxic and solid urban waste dumping (Mazza et al. 2015; Marfe and Di Stefano 2016). At the same time, a number of industrial activities are also present in the

nearby rural areas of the Campania region, where it is also common practice to use biomass resources for the combustion in heating systems. These results confirmed the findings of other studies which pointed out that BbF, BkF and BaP compounds were usually associated with biomass combustion (waste and wood combustion) in most rural areas in the Campania region (Arienzo et al. 2015; Albanese et al. 2015a). Similarly, Bixiong et al. (2006) suggested BbF and

Fig. 9 Interpolated factor score map of the factor 2 (F2). Highest factor scores are correlated with BbF, BkF and BaP compounds, and are displayed in the metropolitan area of Naples (Campania)



BkF as indicators of coal and wood combustions, whilst BaP has been used successfully as marker for biomass combustion (Simcik et al. 1999; Fang et al. 2004). The lower F_2 factor scores (from -2.55 to -1.0) were found in the same areas where F_1 (see Fig. 8) displayed its lower factor score values. These areas were identified in Calabria (e.g. nearby Nucleo Industrial areas, Cosenza), and in areas nearby the oil field in Sicily (Priolo, Ragusa and Gela). These results support the observation that Nap, Acy and Ace compounds are usually related to oil combustion from petroleum exploitation industries (Masclet et al. 1987; Budzinski et al. 1997), which are present both in Basilicata and in Sicily.

Factor 3, contributing to 13.36% of the total variance, is dominated by positively correlated IcdP, DahA and BghiP and antithetic Ant. The factor score map (Fig. 10) presented high values (from 1.04 to 2.42) in and around most urban sites such as Naples, Vibo Valentia and Catania, and around the largest oil refineries in the study area (Milazzo and Val D'Agri). These findings are supported by the fact that IcdP, DahA and BghiP compounds are usually considered as markers of gasoline engine and crude oil combustion sources (Khalili et al. 1995; Larsen and Baker 2003).

Potential soil toxicity

Using Eq. 1, potential toxicity levels of PAHs (in soils) were calculated and reported as TEQ_{BAP} . Results highlighted a significant variation of TEQ_{BAP} values in the five administrative regions studied (Fig. 11). TEQ_{BAP} values in Sicily ranged from 0.49 to 49.29 $ng\ g^{-1}$ and from 0.20 to 7.228 $ng\ g^{-1}$ in urban and rural areas with mean values of 5.56 and 2.04 $ng\ g^{-1}$, respectively, indicating that urban soils have higher TEQ_{BAP} toxicity levels compared to those in rural areas. Similarly, in Calabria and Latium average TEQ_{BAP} values were 3.40 and 23.87 $ng\ g^{-1}$ in urban areas, and 1.79 and 2.12 $ng\ g^{-1}$ in rural areas, respectively. The slightly higher mean value displayed by Latium (TEQ_{BAP} 23.87 $ng\ g^{-1}$) in urban areas can be associated with the prevalence of HWMPAHs in the TEQ computation. These compounds are usually related to vehicular emissions, where Rome (Latium region) is the Italian city with the highest motorisation rate (ISTAT 2012; ISPRA 2012).

Interestingly, Campania and Basilicata presented instead higher mean TEQ_{BAP} values in their rural areas

(661 and 2.71 $ng\ g^{-1}$, respectively), whilst the mean TEQ_{BAP} values for their urban areas were 54.93 and 1.19 $ng\ g^{-1}$, respectively.

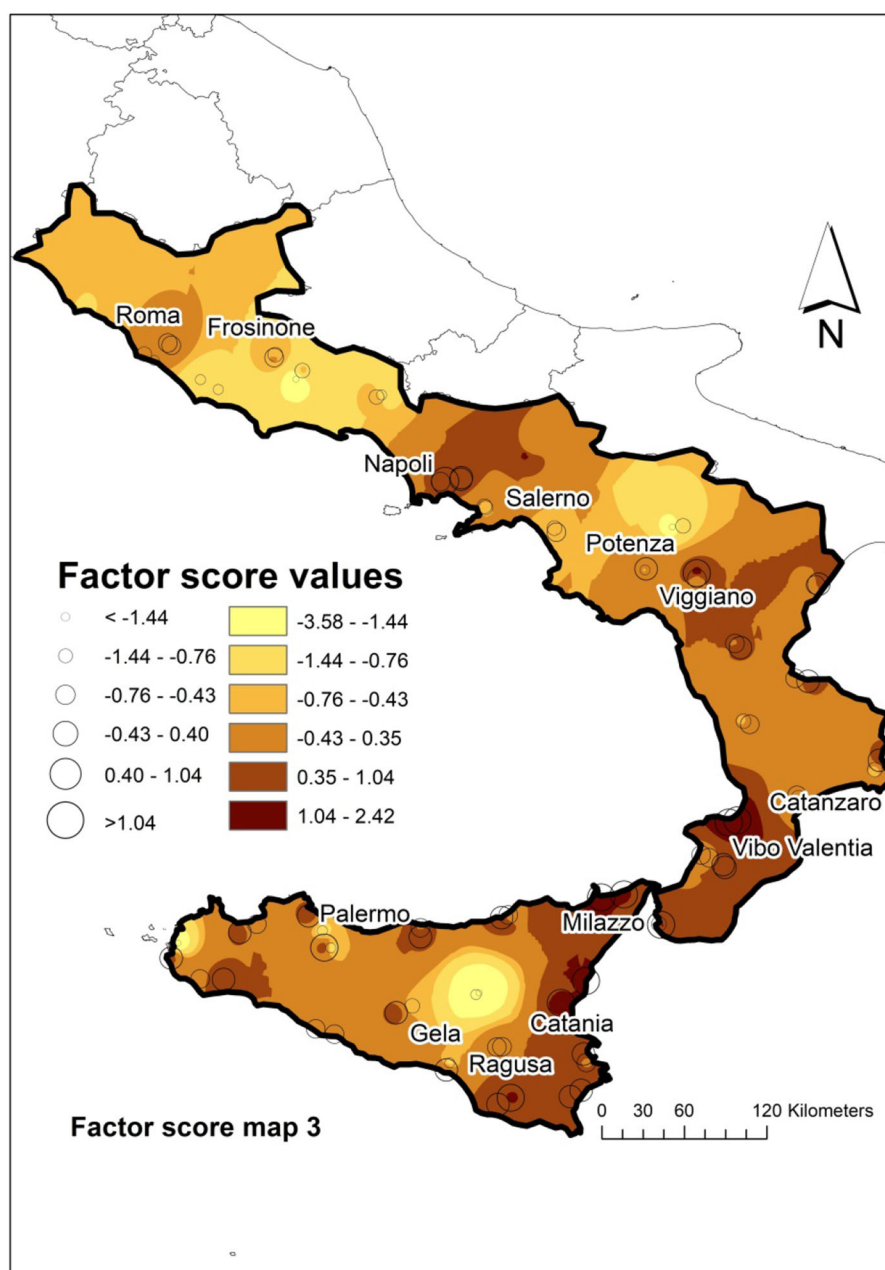
Overall, Campania presented the highest toxicity average values in both its urban and rural areas. The values found in the present study are significant when compared to similar works published worldwide. In urban areas, variability is greater and depends on size of the city and urban fabric. For example, TEQ_{BAP} values found in Gwangju City (Korea) were 13.23 $ng\ g^{-1}$ (Islam et al. 2017), which are comparable to those found in urban areas of Tarragona (Spain) with 64.0 $ng\ g^{-1}$ (Nadal et al. 2007) and the present study, whilst those found in urban soils of a megacity such as Delhi (India) were an order of magnitude higher (218 $ng\ g^{-1}$ —Agarwal et al. 2009), (Table 5). In rural soils studies, values found in the literature vary from 11.20 $ng\ g^{-1}$ in agricultural soils in Poland (Maliszewska-Kordybach et al. 2009), 14.30 $ng\ g^{-1}$ in some soils of Norway (Nam et al. 2008), to values of 83.12 $ng\ g^{-1}$ found in rural soils in the UK (Nam et al. 2008) (Table 5). Several studies (Means et al. 1980; Agapkina et al. 2007; Bhupander et al. 2012) revealed that the occurrence of individual PAHs in soils depends on the land use and the settlement of main PAHs pollution sources. In fact, biomass burning is considered the major source of PAHs in rural soils, and urban soils are likely polluted by release of PAHs from vehicular emission around the heavy traffic roads (Khalili et al. 1995; Morillo et al. 2007; Albanese et al. 2015a). This study unveiled that urban areas of Sicily, Calabria and Latium displayed higher TEQ_{BAP} values than in their rural areas, but rural soils of Basilicata and Campania revealed the contrary.

Among all the findings, it is perhaps arguable that the most striking of all is that related to the metropolitan and rural areas around Naples, characterised by high toxicity levels. These results are partially confirmed by a more detailed study carried out by Albanese et al. (2015a) and by a larger study (known as Campania Trasparente—still in progress), with sampling of soils and air matrices covering the entire regional territory (Qu et al. 2018).

Conclusions

This study carried out a regional survey of urban and rural topsoils (80 samples) in five administrative

Fig. 10 Interpolated factor score map of the third factor (F3). High factor score values corresponding to IcdP, DahA and BghiP variables, are displayed in Naples, Vibo Valentia, Catania and around the largest oil refineries in the study area (Milazzo and Val D'Agri)



regions of the south of Italy, in an attempt to shed light to the main potential sources and patterns of pollution for 16 PAHs compounds (US EPA priority compounds). Measuring molecular PAHs diagnostic ratios allowed to identify and clearly show the main areas of concern, as well as giving an indication of the most likely sources for the PAHs compounds in the various regions. Results strongly pointed to the direction of pyrogenic sources (e.g. vehicular emission, fuels and

biomass combustions) for some areas (urban areas of Latium and rural areas of Campania regions). In particular, these areas were found to have high BAP concentration levels, which in turn indicated high levels of toxicity equivalent quantity (TEQ_{BAP}). The highest level of TEQ (661 ng g^{-1}) was found in metropolitan and rural areas of the Campania region, which could be related to the presence of a large solid waste incinerator plant, illegal waste disposal and

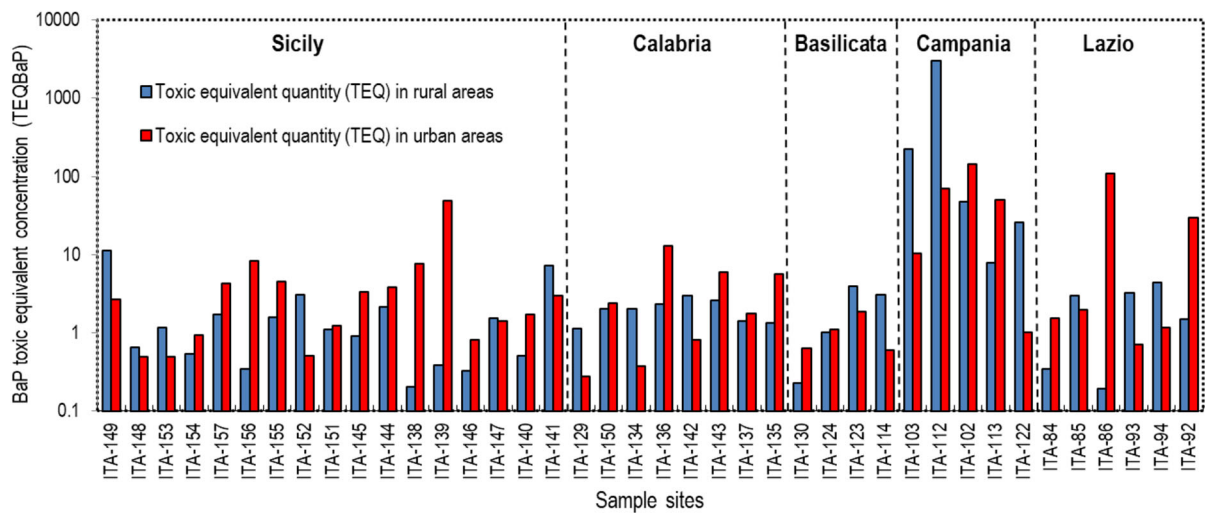


Fig. 11 Variation of the BaP toxicity equivalent quantity (TEQ_{BaP}) values in the survey area; logarithmic scale is applied on the Y axis

Table 5 Means TEQ_{BaP} values in the five studied regions compared to other selected survey studies

Regions	Type of area	TEQ_{BaP} ($ng\ g^{-1}$)	References
Sicily	Urban	5.56	This study
	Rural	2.04	This study
Calabria	Urban	3.40	This study
	Rural	1.79	This study
Basilicata	Urban	1.19	This study
	Rural	2.71	This study
Campania	Urban	54.93	This study
	Rural	661	This study
Latium	Urban	23.87	This study
	Rural	2.12	This study
Gwangju City (Korea)	Urban	14.30	Islam et al. (2017)
Tarragonain (Spain)	Urban	64	Nadal et al. (2007)
Delhi (India)	Urban	218	Agarwal et al. (2006)
Rural soils in Norway	Rural	14.3	Nam et al. (2008)
Agriculture soils in Poland	Agriculture soil	11.9	Maliszewska-Kordybach et al. (2009)
Rural soils in the UK	Rural	83.5	Nam et al. (2008)

burning, and illegal practice of industrial toxic and solid urban waste dumping (both known in the area). Results from this study can represent a fundamental step to understand the distribution, sources and toxicity levels of PAHs in the soils of these regions, giving an impetus to follow up with more detailed surveys. Given the carcinogenic and mutagenic properties of these contaminants, it is envisaged that the

findings from this study will help initiate an assessment of human health risks from to PAH exposures in rural and urban areas in this Mediterranean region.

Acknowledgements This work was supported by grant from the Ministero dell'Università e della Ricerca Scientifica—Industrial Research Project 'BioPolis' PON03PE_00107_01, funded in the frame of Operative National Programme Research and Competitiveness 2007–2013, D. D. Prot. N.713/Ric. 29/10/

2010 and (EnerbioChem) PON01_01966, funded in the frame of the Operative National Programme Research and Competitiveness 2007–2013 D. D. Prot. n. 01/Ric. 18.1.2010.

References

- Agapkina, G. I., Chikov, P. A., Shelepchikov, A. A., et al. (2007). Polycyclic aromatic hydrocarbons in soils of Moscow. *Vest. Moskov University, Series 17: pochvoved* (Vol. 3, pp. 38–46).
- Agarwal, T., Khillare, P. S., & Shridhar, V. (2006). PAHs contamination in bank sediment of the Yamuna River, Delhi, India. *Environmental Monitoring and Assessment*, 123, 151–166.
- Agarwal, T., Khillare, P. S., Shridhar, V., & Ray, S. (2009). Pattern, sources and toxic potential of PAHs in the agricultural soils of Delhi, India. *Journal of hazardous materials*, 163, 1033–1039.
- Aichner, B., Glaser, B., & Zech, W. (2007). Polycyclic aromatic hydrocarbons and polychlorinated biphenyls in urban soils from Kathmandu, Nepal. *Organic Geochemistry*, 38, 700–715.
- Albanese, S., De Vivo, S., Lima, A., & Cicchella, D. (2007). Geochemical background and baseline values of toxic elements in stream sediments of Campania region (Italy). *Journal of Geochemical Exploration*, 93, 21–34.
- Albanese, S., Fontaine, B., Chen, W., Lima, A., Cannatelli, C., Piccolo, A., et al. (2015a). Polycyclic aromatic hydrocarbons in the soils of a densely populated region and associated human health risks: The Campania plain (southern Italy) case study. *Environmental Geochemistry and Health*, 37, 1–20.
- Albanese, S., Sadeghi, M., Lima, A., Cicchella, D., Dinelli, E., Valera, P., et al. (2015b). GEMAS:cobalt, Cr, Cu and Ni distribution in agricultural and grazing land soil of Europe. *Journal of Geochemical Exploration*, 154, 81–93.
- Arienzo, M., Albanese, S., Lima, A., Cannatelli, C., Aliberti, F., Cicotti, F., et al. (2015). Assessment of the concentrations of polycyclic aromatic hydrocarbons and organochlorine pesticides in soils from the Sarno River basin, Italy, and ecotoxicological survey by *Daphnia magna*. *Environmental Monitoring and Assessment*, 187, 52.
- Bevilacqua, M., & Braglia, M. (2002). Environmental efficiency analysis for ENI oil refineries. *Journal of Cleaner Production*, 10(1), 85–92.
- Bhupander, K., Gargi, G., Richa, G., Dev, P., Sanjay, K., & Shekhar, S. C. (2012). Distribution, composition profiles and source identification of polycyclic aromatic hydrocarbons in roadside soil of Delhi, India. *Journal of Environmental Earth Sciences*, 2(1), 10–22.
- Bixiong, Y., Zhihuan, Z., & Ting, M. (2006). Pollution sources identification of polycyclic aromatic hydrocarbons of soils in Tianjin area, China. *Chemosphere*, 64(4), 525–534.
- Bucheli, T. D., Blum, F., Desaulles, A., & Gustaffson, O. (2004). Polycyclic aromatic hydrocarbons, black carbon, and molecular markers in soils of Switzerland. *Chemosphere*, 56, 1061–1076.
- Budzinski, H., Jones, I., Bellocq, J., Pierard, C., & Garrigues, P. (1997). Evaluation of sediment contamination by polycyclic aromatic hydrocarbons in the Gironde estuary. *Marine Chemistry*, 58, 85–97.
- Chen, L., Ran, Y., Xing, B., Mai, B., He, J., Wei, X., et al. (2005). Contents and sources of polycyclic aromatic hydrocarbons and organochlorine pesticides in vegetable soils of Guangzhou, China. *Chemosphere*, 60, 879–890.
- Cheng, Q., Agterberg, F. P., & Ballantyne, S. B. (1994). The separation of geochemical anomalies from background by fractal methods. *Journal of Geochemical Exploration*, 51, 109–130.
- Cheng, Q., Bonham-Carter, G. F., & Raines, G. L. (2001). GeoDAS: A new GIS system for spatial analysis of geochemical data sets for mineral exploration and environmental assessment. In *The 20th international geochemical exploration symposium (IGES)* (pp. 42–43). Santiago de Chile, 6/5–10/5.
- Doglion, C., & Flores, G. (1997). Italy. In E. M. Moores & R. W. Fairbridge (Eds.), *Encyclopedia of European and Asian regional geology* (pp. 414–435). New York: Chapman & Hall.
- Dong, T. T. T., & Lee, B. K. (2009). Characteristics, toxicity, and source apportionment of polycyclic aromatic hydrocarbons (PAHs) in road dust of Ulsan, Korea. *Chemosphere*, 74, 1245–1253.
- DPCM. (1990). *Decreto Presidente del Consiglio dei Ministri (DPCM)*, 30 Novembre 1990 (Italian).
- Environmental Protection Agency (EPA). (1984). *Health effects assessment for polycyclic aromatic hydrocarbons (PAH)*. EPA 540/1-86-013. Environmental Criteria and Assessment Office, Cincinnati, OH.
- ESRI (Environmental Systems Research Institute). (2012). *ArcGIS desktop: Release 10*. Redlands, CA: Environmental Systems Research Institute.
- Fang, G.-C., Chang, K.-C., Lu, C., & Bai, H. (2004). Estimation of PAHs dry deposition and BaP toxic equivalency factors (TEFs) study at Urban, Industry Park and rural sampling sites in central Taiwan, Taichung. *Chemosphere*, 55(6), 787–796. <https://doi.org/10.1016/j.chemosphere.2003.12.012>.
- Filzmoser, P., Hron, K., & Reimann, C. (2009). Principal component analysis for compositional data with outliers. *Environmetrics*, 20(6), 621–632.
- Hwang, H. M., Wade, T. L., & Sericano, J. L. (2003). Concentrations and source characterization of polycyclic aromatic hydrocarbons in pine needles from Korea, Mexico, and United States. *Atmospheric Environment*, 37, 2259–2267.
- IARC (International Agency for Research on Cancer). (1983). *IARC monographs on the evaluation of the carcinogenic risk of chemicals to human. Polynuclear aromatic compounds, Part I, chemical, environmental, and experimental data*. Geneva: World Health Organization.
- Islam, M. N., Park, M., Jo, Y. T., Nguyen, X. P., Park, S. S., Chung, S.-Y., et al. (2017). Distribution, sources, and toxicity assessment of polycyclic aromatic hydrocarbons in surface soils of the Gwangju City, Korea. *Journal of Geochemical Exploration*, 180, 52–60. <https://doi.org/10.1016/j.gexplo.2017.06.009>.

- ISPRA Istituto SperlaPAmbientale. (2012). *Qualità dell'ambiente urbano. VIII Rapporto*. ISPRA-ARPA-APPA, roma (Italian).
- ISPRA. (2014). *Audizione dell'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) presso la Commissione Agricoltura, congiuntamente con la Commissione Ambiente, della Camera sul consumo di suolo*. Audizione, Roma, 27/2/2014 (Italian).
- ISTAT. (2012). *Indicatori ambientali urbani*, Anno 2011. ISTAT, Roma (Italian).
- ISTAT. (2016). *Resident population*. <https://www.istat.it/en/population-andhouseholds>. January 1st, 2017.
- Katsoyiannis, A., Terzi, E., & Cai, Q.-Y. (2007). On the use of PAH molecular diagnostic ratios in sewage sludge for the understanding of the PAH sources. Is this use appropriate? *Chemosphere*, 69, 1337–1339.
- Khalili, N. R., Scheff, P. A., & Holsen, T. M. (1995). PAH source fingerprints for coke ovens, diesel and gasoline engine highway tunnels and wood combustion emissions. *Atmospheric Environment*, 29, 533–542.
- Kozak, K., Ruman, M., Kosek, K., Karasiński, G., Stachnik, L., & Żaneta Polkowska, Z. (2017). Impact of volcanic eruptions on the occurrence of PAHs compounds in the aquatic ecosystem of the Southern Part of West Spitsbergen (Hornsund Fjord, Svalbard). *Water*, 9(1), 42. <https://doi.org/10.3390/w9010042>.
- Larsen, R. K., III, & Baker, J. E. (2003). Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: A comparison of three methods. *Environmental Science and Technology*, 37, 1873–1881.
- Legislative Decree 152/2006 Decreto Legislativo 3 aprile. (2006). n. 152, "Norme in materia ambientale". Gazzetta Ufficiale n. 88 14-4-2006, Suppl Ord n. 96. <http://www.camera.it/parlam/leggi/deleghe/06152dl.htm>.
- Li, W. H., Tian, Y. Z., Shi, G. L., Guo, C. S., Li, X., & Feng, Y. C. (2012). Concentrations and sources of PAHs in surface sediments of the Fenhe reservoir and watershed, China. *Ecotoxicology and Environmental Safety*, 75, 198–206. <https://doi.org/10.1016/j.ecoenv.2011.08.021>.
- Lima, A., De Vivo, B., Cicchella, D., Cortini, M., & Albanese, S. (2003). Multifractal IDW interpolation and fractal filtering method in environmental studies: An application on regional stream sediments of Campania Region (Italy). *Applied Geochemistry*, 18(12), 1853–1865.
- Maliszewska-Kordybach, B., Smreczak, B., & Klimkowicz-Pawlas, A. (2009). Concentrations, sources, and spatial distribution of individual polycyclic aromatic hydrocarbons (PAHs) in agricultural soils in the Eastern part of the EU: Poland as a case study. *Science of the Total Environment*, 407(12), 3746–3753. <https://doi.org/10.1016/j.scitotenv.2009.01.010>.
- Marfè, G., & Di Stefano, C. (2016). The evidence of toxic wastes dumping in Campania, Italy. *Critical Reviews in Oncology/Hematology*, 105, 84–91. <https://doi.org/10.1016/j.critrevonc.2016.05.007>.
- Masclat, P., Bresson, M. A., & Mouvier, G. (1987). Polycyclic aromatic hydrocarbons emitted by power stations, and influence of combustion conditions. *Fuel*, 66, 556–562.
- Mazza, A., Piscitelli, P., Neglia, C., Della Rosa, G., & Iannuzzi, L. (2015). Illegal dumping of toxic waste and its effect on human health in Campania, Italy. *International Journal of Environmental Research and Public Health*, 12, 6818–6831. <https://doi.org/10.3390/ijerph120606818>.
- Means, J. C., Wood, S. G., Hassett, J. J., & Banwart, W. L. (1980). Sorption of polynuclear aromatic hydrocarbons by sediments and soils. *Environmental Science and Technology*, 14(12), 1524–1528.
- Menzie, C. A., Potocki, B. B., & Santodonato, J. (1992). Exposure to carcinogenic PAHs in the environment. *Environmental Science and Technology*, 26(7), 1278–1283.
- Morillo, E., Romero, A. S., Madrid, L., Villaverde, J., & Maqueda, C. (2008). Characterization and sources of PAHs and potentially toxic metals in urban environments of Seville (Southern Spain). *Water Air Soil Pollution*, 187, 41–51.
- Morillo, E., Romero, A. S., Maqueda, C., Madrid, L., Ajmone-Marsan, F., Grcman, H., et al. (2007). Soil pollution by PAHs in urban soils: A comparison of three European cities. *Journal of Environmental Monitoring*, 9, 1001–1008.
- Nadal, M., Schuhmacher, M., & Domingo, J. L. (2004). Levels of PAHs in soil and vegetation samples from Tarragona County, Spain. *Environmental Pollution*, 132, 1–11.
- Nadal, M., Schuhmacher, M., & Domingo, J. L. (2007). Levels of metals, PCBs, PCNs and PAHs in soils of a highly industrialized chemical/petrochemical area: Temporal trend. *Chemosphere*, 66(2), 267–276. <https://doi.org/10.1016/j.chemosphere.2006.05.020>.
- Nam, J. J., Thomas, G. O., Jaward, F. M., Steinnes, E., Gustafsson, O., & Jones, K. C. (2008). PAHs in background soils from Western Europe: Influence of atmospheric deposition and soil organic matter. *Chemosphere*, 70, 1596–1602.
- Nisbet, I. C. T., & LaGoy, P. K. (1992). Toxic equivalency factors (TEFs) for polycyclic aromatic hydrocarbons (PAHs). *Regulatory Toxicology and Pharmacology*, 16, 290–300.
- Ottesen, R. T., Birke, M., Finne, T. E., Gosar, M., Locutura, J., Reimann, C., et al. (2013). Mercury in European agricultural and grazing land soils. *Applied Geochemistry*, 33, 1–12.
- Pandey, P. K., Patel, K. S., & Lenicek, J. (1999). Polycyclic aromatic hydrocarbons: Need for assessment of health risks in India? Study of an urban-industrial location in India. *Environmental Monitoring and Assessment*, 59, 287–319.
- Pawlowsky-Glahn, V., & Buccianti, A. (2011). *Compositional data analysis: Theory and applications* (p. 740). Hoboken: Wiley.
- Qu, C., Doherty, A. L., Xing, X., Sun, W., Albanese, S., Lima, A., et al. (2018). Chapter 20—Polyurethane foam-based passive air samplers in monitoring persistent organic pollutants: Theory and application. In B. De Vivo, A. Lima, H. E. Belkin (Eds.), *Environmental geochemistry* (2nd ed., pp. 521–542). Elsevier. <https://doi.org/10.1016/B978-0-444-63763-5.00021-5>.
- Ravindra, K., Wauters, E., & Van Grieken, R. (2008). Variation in particulate PAHs levels and their relation with the transboundary movement of the air masses. *Science of the Total Environment*, 396, 100–110.
- Reimann, C., Filzmoser, P., & Garrett, R. (2002). Factor analysis applied to regional geochemical data: Problems and possibilities. *Applied Geochemistry*, 17(3), 185–206.

- Reimann, C., Filzmoser, P., Garrett, R.G., & Dutter, R., (2008). Statistical data analysis explained. In *Applied environmental statistics with R* (p. 362). Chichester: Wiley.
- Reimann, C., Flem, B., Fabian, K., Birke, M., Ladenberger, A., Négrel, P., et al. (2012). Lead and lead isotopes in agricultural soils of Europe: The continental perspective. *Applied Geochemistry*, 27, 532–542.
- Reimann, C., Garrett, R. G., & Filzmoser, P. (2005). Background and threshold—Critical comparison of methods of determination. *Science of the Total Environment*, 346, 1–16.
- Simcik, M. F., Eisenreich, S. J., & Lioy, P. J. (1999). Source apportionment and source/sink relationships of PAHs in the coastal atmosphere of Chicago and Lake Michigan. *Atmospheric Environment*, 33(30), 5071–5079. [https://doi.org/10.1016/S1352-2310\(99\)00233-2](https://doi.org/10.1016/S1352-2310(99)00233-2).
- Soclo, H. H., Garrigues, P., & Ewald, M. (2000). Origin of polycyclic aromatic hydrocarbons (PAHs) in coastal marine sediments: Case studies in Cotonou (Benin) and Aquitaine (France) areas. *Marine Pollution Bulletin*, 40(5), 387–396. [https://doi.org/10.1016/S0025-326X\(99\)00200-3](https://doi.org/10.1016/S0025-326X(99)00200-3).
- Sofowote, U. M., McCarry, B. E., & Marvin, C. H. (2008). Source apportionment of PAH in Hamilton Harbour suspended sediments: Comparison of two factor analysis methods. *Environmental Science and Technology*, 42(60), 07–14.
- Spaziani, F., Angelone, M., Coletta, A., Salluzzo, A., & Cremisini, C. (2008). Determination of platinum group elements and evaluation of their traffic-related distribution in Italian urban environments. *Analytical Letters*, 41(14), 2658–2683. <https://doi.org/10.1080/00032710802363503>.
- Tang, L., Tang, X., Zhu, Y., Zheng, M., & Miao, Q. (2005). Contamination of polycyclic aromatic hydrocarbons (PAHs) in urban soils in Beijing, China. *Environment International*, 31(6), 822–828. <https://doi.org/10.1016/j.envint.2005.05.031>.
- Templ, M., Hron, K., & Filzmoser, P., (2011). *Robcompositions: Robust estimation for compositional data*. Manual and package, Version 1.4.4.
- Thiombane, M., Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Cavaliere, M., et al. (2017). Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C–A fractal model. *Journal of Geochemical Exploration*, 189, 85–99.
- Tobiszewski, M., & Namieśnik, J. (2012). PAH diagnostic ratios for the identification of pollution emission sources. *Environmental Pollution*, 162, 110–119.
- US EPA (2002). *US Environmental Protection Agency Guidance on environmental data verification and data validation* (p. 96). EPA QA/G-8. Washington, DC: Office of Environmental Information.
- Vane, C. H., Kim, A. W., Beriro, D. J., Cave, M. R., Knights, K., Moss-Hayes, V., et al. (2014). Polycyclic aromatic hydrocarbons (PAH) and polychlorinated biphenyls (PCB) in urban soils of Greater London, UK. *Applied Geochemistry*, 51, 303–314.
- Yunker, M. B., Macdonald, R. W., Vingarzan, R., Mitchell, R. H., Goyette, D., & Sylvestre, S. (2002). PAHs in the Fraser river basin: A critical appraisal of PAH ratios as indicators of PAH source and composition. *Organic Geochemistry*, 33, 489–515.
- Zhang, H. B., Luo, Y. M., Wong, M. H., Zhao, Q. G., & Zhang, G. L. (2006). Distributions and concentrations of PAHs in Hong Kong soils. *Environmental Pollution*, 141, 107–114.

Paper 7

Status, sources and contamination levels of organochlorine pesticides residues in urban and agricultural areas: A preliminary review in central-southern Italian soils

Matar Thiombane, Attila Petrik, Marcello Di Bonito, Stefano Albanese, Daniela Zuzolo, Domenico Cicchella, Annamaria Lima, Chengkai Qu, Shihua Qi, Bendetto De Vivo

Journal of Environmental Science and Pollution Research, September 2018, Volume 25, Issue 26, pages 361–382



Status, sources and contamination levels of organochlorine pesticide residues in urban and agricultural areas: a preliminary review in central–southern Italian soils

Matar Thiombane¹ · Attila Petrik¹ · Marcello Di Bonito² · Stefano Albanese¹ · Daniela Zuzolo³ · Domenico Cicchella³ · Annamaria Lima¹ · Chengkai Qu¹ · Shihua Qi⁴ · Benedetto De Vivo^{5,6}

Received: 18 March 2018 / Accepted: 2 July 2018 / Published online: 6 July 2018

© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Organochlorine pesticides (OCPs) are synthetic chemicals commonly used in agricultural activities to kill pests and are persistent organic pollutants (POPs). They can be detected in different environmental media, but soil is considered an important reservoir due to its retention capacity. Many different types of OCPs exist, which can have different origins and pathways in the environment. It is therefore important to study their distribution and behaviour in the environment, starting to build a picture of the potential human health risk in different contexts. This study aimed at investigating the regional distribution, possible sources and contamination levels of 24 OCP compounds in urban and rural soils from central and southern Italy. One hundred and forty-eight topsoil samples (0–20 cm top layer) from 78 urban and 70 rural areas in 11 administrative regions were collected and analysed by gas chromatography–electron capture detector (GC–ECD). Total OCP residues in soils ranged from nd (no detected) to 1043 ng/g with a mean of 29.91 ng/g and from nd to 1914 ng/g with a mean of 60.16 ng/g in urban and rural area, respectively. Endosulfan was the prevailing OCP in urban areas, followed by DDTs, Drins, Methoxychlor, HCHs, Chlordane-related compounds and HCB. In rural areas, the order of concentrations was Drins > DDTs > Methoxychlor > Endosulfans > HCHs > Chlordanes > HCB. Diagnostic ratios and robust multivariate analyses revealed that DDT in soils could be related to historical application, whilst (illegal) use of technical DDT or dicofol may still occur in some urban areas. HCH residues could be related to both historical use and recent application, whilst there was evidence that modest (yet significant) application of commercial technical HCH may still be happening in urban areas. Drins and Chlordane compounds appeared to be mostly related to historical

Highlights • High concentration of banned organochlorine pesticides in urban and rural soils

- OCP residues in soil are likely related to both historical and recent applications
- DDT and HCH metabolites in soils are mainly related to historical applications
- Slight illegal fresh use of technical DDT and HCH residues mostly in urban soils
- High contamination levels are revealed in agricultural soils of Campania and Apulia

Responsible editor: Hongwen Sun

✉ Matar Thiombane
thiombane.matar@unina.it

¹ Department of Earth, Environment and Resources Sciences (DiSTAR), University of Naples “Federico II,” Complesso Universitario di Monte Sant’ Angelo, Via Cintia snc, 80126 Naples, Italy

² School of Animal, Rural and Environmental Sciences, Nottingham Trent University, Brackenhurst Campus, Southwell NG25 0QF, UK

³ Department of Science and Technology, University of Sannio, via dei Mulini 59/A, 82100 Benevento, Italy

⁴ State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, 430074 Wuhan, People’s Republic of China

⁵ Pegaso University, Piazza Trieste e Trento 48, 80132 Naples, Italy

⁶ Benecon Scarl, Dip. Ambiente e Territorio, Via S. Maria di Costantinopoli 104, 80138 Naples, Italy

application, whilst Endosulfan presented a complex mix of results, indicating mainly historical origin in rural areas as well as potential recent applications on urban areas. Contamination levels were quantified by Soil Quality Index (SoQI), identifying high levels in rural areas of Campania and Apulia, possibly due to the intensive nature of some agricultural practices in those regions (e.g., vineyards and olive plantations). The results from this study (which is in progress in the remaining regions of Italy) will provide an invaluable baseline for OCP distribution in Italy and a powerful argument for follow-up studies in contaminated areas. It is also hoped that similar studies will eventually constitute enough evidence to push towards an institutional response for more adequate regulation as well as a full ratification of the Stockholm Convention.

Keywords Organochlorine pesticides · Italian soils · Diagnostic ratios · Multivariate analysis · Soil quality index · Contamination

Introduction

The Stockholm Convention (2005) banned the use of Persistent Organic Pollutants (POPs), with the aim of protecting human health and the environment. The initial list prepared in 2003 included Aldrin, Dieldrin, Endrin, Chlordane, Heptachlor, Hexachlorobenzene (HCB) and Dichlorodiphenyltrichloroethane (DDT). This list was then expanded with other potential POPs in 2011 (Stockholm Convention on Persistent Organic Pollutants 2011): Hexachlorocyclohexane (HCH, including Lindane), Methoxychlor, and Endosulfan (Stockholm Convention 2005, 2011). These organic pollutants are considered long-range transport compounds based on their ubiquity, persistence and bioaccumulation potential in different environmental media (Weinberg 1998; Szeto and Price 1991; Fang et al. 2017), as well as their high toxicity to humans and non-target organisms (WHO 2003; Nizzetto et al. 2006; Moeckel et al. 2008; Kim et al. 2017). They sink in different environmental matrices, such as air, water and soils, and further accumulate in the food chain (Prapamontol and Stevenson 1991; Suchan et al. 2004; Qu et al. 2016). Soil continues to be a potential medium of exposure of OCPs, and its biofilms and physico-chemical properties may influence fate and behaviours of OCP metabolites through different degradation phenomena (Weinberg 1998; WHO 2003).

Since agricultural practices are a very important economic resource for Italy, this makes it the third OCP user among European Union countries (Eurostat 2014). In Italy, OCPs are used in most agricultural activities, in forestry as well as ornamental plants in urban garden preservation against insects, fungal or animal pests. It is well-known that Italy is the only European Union (EU) country that has not ratified the Stockholm Conventions, though the production and use of Aldrin, Chlordane, Dieldrin, Endrin, DDT, Heptachlor, HCB and HCH in its territory have been strictly restricted in harmony with several other regulatory schemes via the Rotterdam Convention in 1998, the European Directive in 2000

(Persistent organic pollutants amending Directive 79/117/EEC) and the United Nations Economic Commission for Europe POPs Protocol (UNECE 2010). Moreover, Italian environmental law (D. Lgs. 152/2006) established guideline threshold values that regulate the mitigation of OCPs in soils. This regulation guided a recent evaluation of the levels (for DDT) and residues (for HCH) in the Campania plain (Arienzo et al. 2015; Qu et al. 2016) and in agricultural soils in the province of Latina (Latium) (Donnarumma et al. 2009). However, this legislation does not involve OCPs such as Endosulfan and Methoxychlor: these compounds have been associated to both environmental and human health risks due to concerns that they are carcinogen, teratogen and male reproductive toxicants (PANNA 2008; US EPA 2007; Silva and Carr 2009; Jayaraj et al. 2016). Whilst recent studies have started to investigate and define the level of OCPs in Italian soils (e.g., Donnarumma et al. 2009; Arienzo et al. 2015; Qu et al. 2016, 2017), there has been no systematic attempt to evaluate their wider distribution and variations across rural and urban areas in Italy. The aim of this study was to begin to establish a regional (and eventually national) baseline based on a large survey carried out in 11 regions of central and southern Italy. The main objectives of this study were as follows:

- (1) to identify the regional distribution of OCP compounds in Italian soils;
- (2) to evaluate their potential sources by using OCP diagnostic ratios as well as robust compositional biplot and factor analysis; and
- (3) to quantify OCP contamination levels by using Soil Quality index (SoQI) in urban and rural soils.

This study is important because it will represent a fundamental stepping stone to build a long-overdue national picture of OCP status in Italy. It is envisaged that the results of this study should trigger more detailed surveys in contaminated areas as well as ad hoc risk-based studies, which in the long term will constitute a strong-enough argument to cause an adequate institutional response by the Italian regulating authorities.

Materials and methods

Study area

The survey area included 4 administrative regions (Latium, Marche, Tuscany and Umbria) from central and 7 (Abruzzo, Apulia, Basilicata, Calabria, Campania, Molise and Sicily) from southern Italy (Fig. 1).

The total survey area (considering administrative regional boundaries) extended to 157,716 km² with 31.26 million of

inhabitants, mostly grouped in main urban areas (ISTAT 2016). Most of the land is used for agricultural and forestry activities. Agriculture occupies one fourth of the land available, which includes cultivation of hilly areas where agriculture results in modifying the natural landscape and resources through terracing, irrigation and soil management (Corona et al. 2012; ISTAT 2013; ISPRA 2014a). Favourable meteorological conditions, dominated by a Mediterranean climate, allow intensive agriculture activities such as vineyards and olive plantations—mostly in hilly areas—as well as

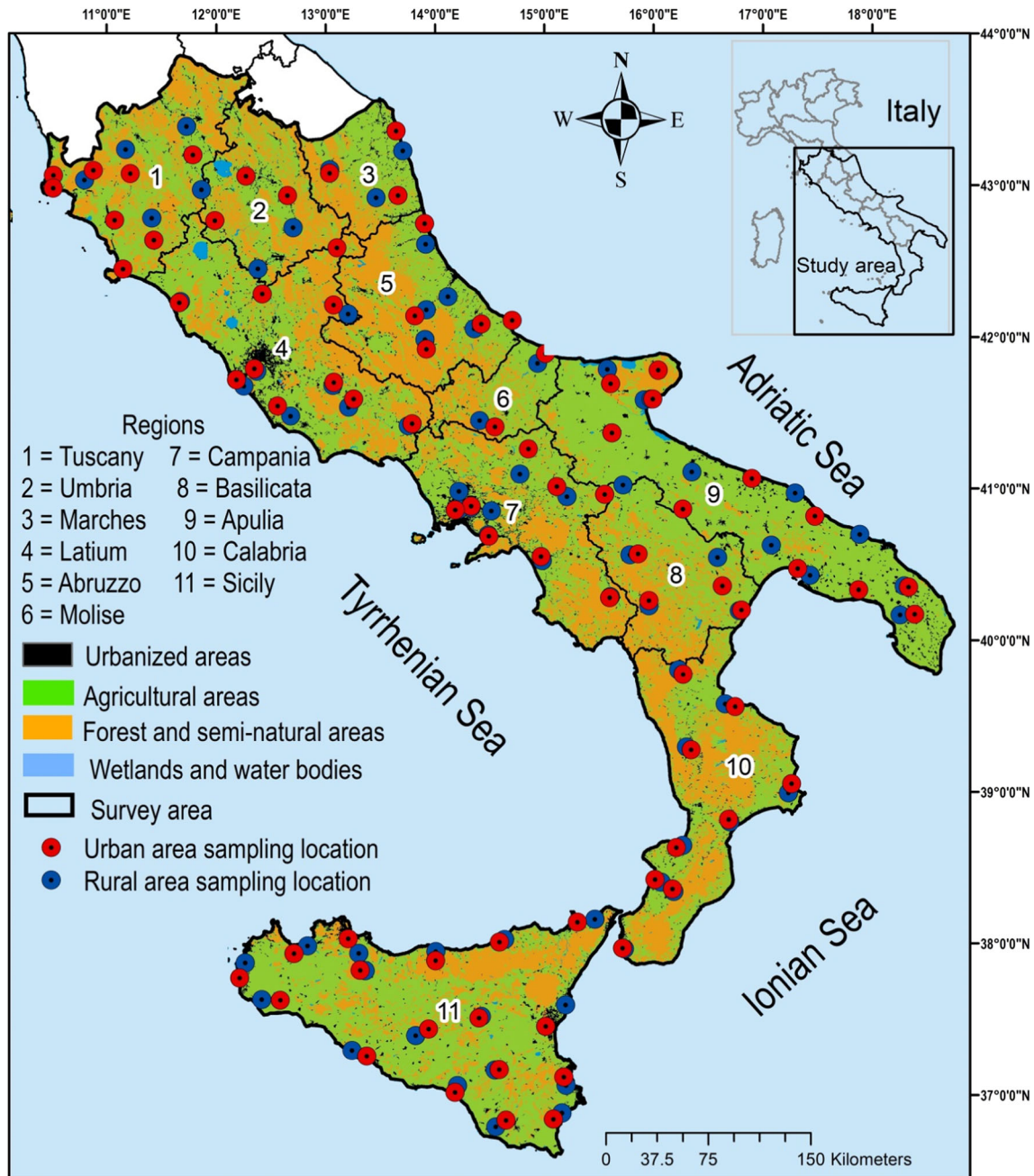


Fig. 1 Land use/land cover of the study area (simplified from Corine Land Cover 2012). Urban (red dots) and rural (blue dots) sampling site locations are displayed

greenhouse production (tomatoes, potatoes, aubergines, peppers, peas and citrus fruits) in coastal areas in Campania, Apulia and Sicily (Costantini and Dazzi 2013). The predominant crops in inland territory are seasonal ones like wheat, maize, potatoes, rice and sugar beet. Most of the forestry lands are composed of broad-leaved trees, with conifers and chestnut making up about one fifth of the total (ISTAT 2013).

Large urban areas, such as Rome (Latium), Naples (Campania), Bari (Apulia) and Palermo (Sicily), are densely populated and surrounded by metropolitan areas where industrial activities, manufactories and intensive agriculture occur (ISTAT 2016).

Soil sampling procedure and preparation

The sampling campaign took place from early April to end of September 2016, with the aim to select the most representative topsoil samples in urban and rural areas throughout 11 regions (Latium, Marches, Tuscany, Umbria, Abruzzo, Apulia, Basilicata, Calabria, Campania, Molise and Sicily) from the centre to southern Italy. In each region, the main urban areas and the nearest rural areas where most of the land is devoted to agricultural activities, were selected. Site selection was carried out by interpreting, using Geographical Information Systems (ArcGIS 2012), information on land use/land cover of the study area (ISPRA 2014b; Corine Land Cover 2012) together with satellite imagery (Google Earth® professional 2016). A total of 148 soil samples were collected with a nominal density of 2 samples/2500 km² (in urban and rural areas) (Fig. 1). Samples have been collected from public gardens in urban areas and from agricultural land (farmlands/cropland) in rural areas. All the samples were collected using a stainless steel scoop, kept in labelled glass bottles and directly stored in ice boxes to minimize the losses caused by volatilization and initial degradation of the organic compounds. Each topsoil sample (from 0 to 20 cm) was made by homogenizing 5 sub-samples at the corners and the centre of a 100 m² square, collecting approximately 1.5 kg in total. The sampling procedure followed the Geochemical Mapping of Agricultural and Grazing Land Soil (GEMAS) sampling procedure described by Reimann et al. (2014). Soil samples were homogenized and sieved using a < 2-mm mesh sieve after removing stones, detritus and residual roots. Finally, composite samples were stored at − 4 °C in the environmental geochemistry laboratory of the University of Naples Federico II (Italy) until instrumental analysis. Geographical coordinates were recorded by geospatial positioning systems (WGS84, GPS) at each sample site.

Extraction procedure and analysis of OCPs

Analyses were carried out by an Agilent 7890A gas chromatograph with a 63Ni electron capture detector (GC-ECD)

equipped with a DB-5 capillary column (30.0 m length, 0.32 mm diameter, 0.25 mm film thickness), in the Key Laboratory of Biogeology and Environmental Geology of Ministry of Education at the University of Geosciences in Wuhan, China (Yang et al. 2008; Qu et al. 2016).

Gas chromatography–mass spectrometry (GC–MS) and gas chromatography–electron capture detector (GC–ECD) are the most common and appropriate systems to investigate organic contaminants in different environmental media. Many authors (Aramendia et al. 2007; Alves et al. 2012) showed the high sensitivity of GC–ECD for organophosphorus and organochlorine pesticides. In this study, the rationale of working with GC–ECD analyser was based on the excellent sensitivity and satisfactory quantification limits, allowing the identification and quantification of pesticides at low levels. A 10 g of dried soil samples was spiked with 20 ng of 2,4,5,6-tetrachloro-m-xylene (TCmX) and decachlorobiphenyl (PCB209) as recovery surrogates and were Soxhlet-extracted with dichloromethane for 24 h. Activated copper granules were added to the collection flask to remove elemental sulphur. The extraction of OCPs was concentrated and solvent-exchanged to n-hexane and further reduced to 2–3 mL by rotary evaporation. The alumina/silica (1:2) gel column (450 °C muffle drying for 4 h, both deactivated with 3% water) was used to purify the extract, and OCPs were eluted with 30 mL of dichloromethane/hexane (2/3). Then, the eluate was concentrated to 0.2 mL under a gentle nitrogen stream, and a known quantity of penta-chloronitrobenzene (PCNB) was added as an internal standard prior to GC–ECD analysis.

Nitrogen was used as carrier gas at 2.5 mL/min under constant-flow mode. Injector and detector temperatures were maintained at 290 and 300 °C, respectively. The oven temperature started from 100 °C (with an equilibration time of 1 min) and rose to 200 °C at a rate of 4 °C/min, then to 230 °C at 2 °C/min and finally reached 280 °C at 8 °C/min, and it was held for 15 min. Two microlitres of each sample was injected into the GC–μECD system for the analysis. Concentration of the individual target OCPs was identified by comparison of their retention times (previously confirmed with GC/MS) and quantified using an internal standard. The GC–MS parameters of the Agilent 6890GC-5975MSD system were the same as those of the Agilent 6890 GC equipped with 63Ni micro-electron capture detector (GC–μECD). The mass spectrometer (MS) was operated in electron impact ionization mode with electron energy of 70 eV. The ion source, quadruple and transfer line temperatures were held at 230, 150 and 280 °C, respectively. Target compounds were monitored in selected ion monitoring (SIM) mode.

Procedure types used for quality assurance and quality control (QA/QC) were as follows: method blank control (procedural blank samples), parallel sample control (duplicate samples), solvent blank control and basic matter control (US EPA 2000). The spiked samples containing internal standard

compounds were analysed simultaneously with soil samples. A procedural blank and a replicate sample were run with every set of 12 samples analysed to check for contamination from solvents and glassware. The limits of detection (LODs) were based on 3:1 S/N ratio. TCmX and PCB 209 were spiked as surrogate standards to judge procedural performance. The surrogate recoveries for TCmX and PCB 209 were 77.8 ± 19.0 and $89.3 \pm 20.3\%$, respectively. The relative standard deviation (RSD) was less than 10%. All OCP concentrations were expressed on an air-dried weight basis.

Geostatistical and multivariate analyses

OCP associations and possible sources were identified by univariate and multivariate statistical analyses as well as diagnostic ratios, compositional biplot and robust factor analysis. Compositional biplot and robust factor analysis allowed to minimize and/or eliminate the presence of outliers and spurious correlation (Pawlowsky-Glahn and Buccianti 2011; Filzmoser et al. 2012). DDT and HCH compounds were chosen for the multivariate computation both for their high toxicity levels and for their proven predominance in Italian soils and air (e.g., Estellano et al. 2012; Pozo et al. 2016; Qu et al. 2016). Biplot statistical analysis (Gabriel 1971) was used to display both samples and variables of the data matrix in terms of the resulting scores and loading (Pison et al. 2003; Otero et al. 2005). For a full description of compositional biplot, several examples are available in the literature (e.g., Maronna et al. 2006; Filzmoser and Hron 2008; Filzmoser et al. 2009; Hron et al. 2010; Thiombane et al. 2018). Factor analysis (FA) was used to explain the correlation structure of the variables through a reduced number of factors (Reimann et al. 2002). This has been successfully employed to evaluate the potential origins of the compounds in relation to their main hypothetical sources (Reimann et al. 2002; Jiang et al. 2009). Isometric logratio (ilr) transformation was applied on raw data prior to multivariate analysis (Filzmoser et al. 2009). R-mode factor analysis was also performed, and the different factors obtained were studied and interpreted in accordance with their presumed sources (Reimann et al. 2002; Albanese et al. 2007).

Two main open-source R packages for statistical software were used: “Compositions” (Van Den Boogaart et al. 2011) and “Robcompositions” (Templ et al. 2011). OCP concentrations and factor score values were mapped for image pattern recognition using GeoDAS (Cheng et al. 2001) and ArcGIS (ESRI 2012) software. GeoDASTM was used to produce dots and interpolated geochemical maps using the multifractal inverse distance weighted (MIDW) algorithm (Cheng et al. 1994; Lima et al. 2003). The concentration–area (C–A) fractal method was applied to classify OCP concentration and factor score ranges in interpolated images.

Assessment of contamination level

Assessment of contaminated sites is a preliminary requirement to reveal potential impact of OCP pesticides on public and ecosystem health (US EPA 1991; CCME 1992; DoE 1995; APAT 2008; DEFRA 2011). The “Soil Quality Index” (SoQI) elaborated by the Canadian Soil Quality Guidelines for Protection of Environment and Human Health Agency (CCME 2007) was implemented to define, classify and prioritize contamination level for each region. Advantages of the SoQI include that it is a robust computation based on three factors for its calculations, namely the following: (1) scope (% of contaminants that do not meet their respective guidelines), (2) frequency (% of individual tests of contaminants that do not meet their respective guidelines) and (3) amplitude (the amount by which the contaminants do not meet their respective guidelines) and it is relatively simple to use. The SoQI was computed using threshold values for residential areas established by the Italian environmental law (D. Lgs. 152/2006) (Table 1) as reference guidelines.

SoQI index provides a quantitative index based on the amalgamation of the three factors (F_1 , F_2 and F_3):

$$F1 = \frac{\sum fx}{\sum Cx} \times 100 \quad (1)$$

$F1$ (scope) represents the percentage of contaminants that do not meet their respective guideline values, where fx is the number of failed contaminants and Cx is the total number of contaminants

$$F2 = \frac{\sum ft_x}{\sum tx} \times 100 \quad (2)$$

$F2$ (frequency) corresponds to the percentage of individual tests that do not meet their respective guidelines values, where

Table 1 Organochlorine pesticide guideline threshold values in soils, fixed by the Italian environmental law (D. Lgs. 152/2006) in residential areas (and/or park areas) and industrial (or/or commercial) sites

	Residential or recreation or park areas sites (ng/g)	Industrial or commercial sites (ng/g)
Aldrin	10	100
α -HCH	10	100
β -HCH	10	500
γ -HCH or Lindane	10	500
δ -HCH	10	100
Chlordane	10	100
DDT, DDE, DDD	10	100
Dieldrin	10	100
Endrin	10	2000

fix represents the number of failed tests and tx symbolizes the number of tests.

$$Ex_i = \frac{Z_{t_i}}{Gv_i} - 1 \quad (3)$$

Ex_i or Excursion is the magnitude by which the contaminant is over/below the respective guideline value. This is calculated as a ratio of the failed test value (Z_{t_i}) and its respective guideline value (Gv_i)

$$Ase = \frac{\sum_{i=1}^n Ex_i}{\sum fix} \quad (4)$$

The average amount by which individual tests are out of compliance corresponds to Ase.

$$F3 = \frac{Ase}{0.01Ase + 0.01} \quad (5)$$

F3 or amplitude represents the amount by which failed test values do not meet their guidelines.

$$SoQI = 100 - \frac{\sqrt{F1^2 + F2^2 + F3^2}}{1.732} \quad (6)$$

Finally, SoQI is calculated by taking the square root of the sum of squared factors divided by 1.732 and extracting it from 100. The 1.732 normalises the SoQI to a range between 0 and 100. The proposed classes are as follows: very low contamination (90–100), low contamination (70–90), medium contamination (50–70), high (30–50) and very high contamination (0–30).

Results and discussion

Residues and pollution sources of OCPs

OCPs

Total OCP residues in soils ranged from “no detected” (nd) to 1043.98 ng/g with a mean of 29.91 ng/g and from nd to 1914.1 ng/g with a mean of 60.16 ng/g in urban and rural area, respectively (Table 2).

The coefficient of variation (CV) ranged from 0.27 to 8.72 and from 1.87 to 6.47 in urban and rural areas, respectively, reflecting a significant spatial variation.

Endosulfan was the most dominant group accounting for 44.42% of the total OCPs, followed by DDTs with 17.60%, Drins (15.75%), Methoxychlor (12.17%), HCHs (6.08%), Chlordane-related compounds (3.53%) and HCB (0.55%) in urban areas (Fig. 2). In agricultural areas, abundances were in the following order: Drins (39.46%) > DDTs (29.94%) >

Methoxychlor (18.22%) > Endosulfan (5.12%) > HCHs (5.06%) > Chlordanes (1.40%) > HCB (0.79%).

Total DDT and derived metabolites

The total concentration of DDTs ranged from nd to 56.97 ng/g (mean = 5.26 ng/g—urban) and from nd to 632.95 ng/g (mean = 18.01 ng/g—rural). The highest DDT concentrations in the urban area, ranging from 24.82 to 56.97 ng/g, were found in the Sarno Basin (Campania), Apulia (Bari and Foggia) and Abruzzo (Fig. 3a). In contrast, the highest DDT concentrations in the rural areas, ranging from 400 to 628 ng/g, were found around Naples (Campania) where the vast majority of intensive agricultural land is located (Fig. 3b). In particular, total DDT concentration presented significantly skewed distributions as well as clear “outliers” (Fig. 3c,d). The latter, observed in rural areas around Naples (Campania—Fig. 3d), can be considered as anomaly concentrations, which could be linked to the input of DDT through agricultural activities. Campania and Apulia are well-known for their large vineyards and olive plantations on their hills and along coastal areas (Costantini and Dazzi 2013; ISPRA 2014a), and high DDT residues may originate from agricultural activities in these areas. As a general observation, urban areas for this study showed lower DDT residues compared to those reported in similar studies, such as that on Beijing urban park soils (Li et al. 2008). On the other hand, some rural areas revealed much higher DDT residues compared to those reported in counterpart studies (Table 3).

Technical DDT is made up of six congener compounds, such as p,p'-DDT, o,p'-DDT, p,p'-DDE, o,p'-DDE, p,p'-DDD and o,p'-DDD. Moreover, it contains 65–80% of p,p'-DDT, 15–21% of o,p'-DDT, up to 4% of p,p'-DDD and impurities (Metcalf 1995). In nature, p,p'-DDE and p,p'-DDD are the two main products of dechlorination of p,p'-DDT by microorganisms and/or physico-chemical properties of soil (Pfaender and Alexander 1972; Mackay et al. 1992). More recently, dicofol has been introduced, which is structurally similar to DDT and contains high impurity of DDT-related compounds (25% of o,p'-DDT) (Qiu et al. 2005). The ratios between the parent compound and its metabolite can provide useful information on the DDT sources. For example, a survey on the formulated dicofol in China found that the ratio of o,p'-DDT/p,p'-DDT in air (Qiu et al. 2005) and soil (Yang et al. 2008) was as high as 7.

In this study, of the various compounds, the p,p'-DDT isomer was predominant, with 34.44% (urban) and 49.43% (rural). Its ranges went from nd to 16.98 ng/g (urban) and from nd to 418 ng/g (rural) (Table 2). The p,p'-DDE isomer had the second highest percentage (34.38%) and ranged from nd to 38.58 ng/g (mean = 1.81 ng/g). This was followed by o,p'-DDD (13.18%), o,p'-DDT (6.39%), o,p'-DDE (6.36%) and p,p'-DDD (5.24%) in urban areas. On the other hand,

Table 2 Descriptive statistics of the 24 OCP compounds (ng/g) in 148 topsoil samples from urban and rural areas of central and southern Italy; min, max and CV indicate the minimum, maximum and coefficient of variation of the dataset, respectively

Compounds (ng/g)	DL	Urban				Rural			
		Min.	Mean	Max.	CV	Min.	Mean	Max.	CV
α -HCH	0.011	n.d	0.22	4.43	2.40	n.d	0.57	19.21	4.15
β -HCH	0.006	n.d	1.10	5.50	1.36	n.d	1.47	20.38	1.87
γ -HCH	0.011	n.d	0.39	14.19	4.19	n.d	0.65	11.29	2.73
δ -HCH	0.01	n.d	0.11	2.72	3.06	n.d	0.36	18.18	6.05
HCHs	–	n.d	1.82	25.08	1.78	n.d	3.05	47.27	2.30
p,p'-DDT	0.025	n.d	1.81	16.99	1.73	n.d	8.90	418.46	5.68
o,p'-DDT	0.02	n.d	0.34	5.04	2.24	n.d	1.68	48.27	4.80
p,p'-DDE	0.019	n.d	1.81	38.59	3.20	n.d	5.40	139.93	4.19
o,p'-DDE	0.021	n.d	0.33	7.56	2.99	n.d	0.27	4.34	2.26
p,p'-DDD	0.006	n.d	0.28	3.05	2.19	n.d	1.11	36.22	4.24
o,p'-DDD	0.025	n.d	0.34	5.04	2.24	n.d	1.68	48.27	4.80
DDTs	–	n.d	5.26	56.98	1.90	n.d	18.01	632.95	4.59
cis-Chlordane	0.018	n.d	0.08	1.77	2.93	n.d	0.11	3.40	4.07
trans-Chlordane	0.021	n.d	0.20	7.71	4.47	n.d	0.10	1.97	3.33
Heptachlor	0.021	n.d	0.15	3.72	3.41	n.d	0.07	1.20	2.33
Heptachlor-epoxide	0.014	n.d	0.62	10.95	2.64	n.d	0.57	13.73	3.43
Aldrin	0.046	n.d	0.25	2.69	1.75	n.d	0.59	13.37	3.27
Dieldrin	0.036	n.d	0.13	1.16	1.75	n.d	0.24	9.80	4.87
Endrin	0.030	n.d	0.36	10.70	3.49	n.d	0.30	6.65	2.94
Endrin aldehyde	0.030	n.d	0.19	1.30	1.26	n.d	0.36	4.79	2.42
Endrin ketone	0.032	n.d	3.78	82.17	2.95	n.d	22.24	1199.97	6.47
α -Endosulfan	0.017	n.d	9.22	710.34	8.72	n.d	0.12	3.02	3.25
β -Endosulfan	0.017	n.d	2.57	176.79	7.81	n.d	0.36	10.36	4.18
Endosulfan-sulphate	0.064	n.d	1.46	17.43	1.92	n.d	2.61	79.74	3.84
HCB	0.009	0.011	0.17	2.39	2.25	n.d	0.48	13.37	3.53
Methoxychlor	0.025	n.d	3.64	53.23	2.23	n.d	10.96	521.79	5.80
OCPs	–	0.0011	29.91	1043.98	0.27	n.d	60.16	1914.10	4.02

agricultural areas presented a higher dominance of p,p'-DDT (9.32%) > p,p'-DDD (6.17%) > (o,p'-DDD (3.64%) > o,p'-DDE (1.49%) followed by p,p'-DDE (29.96%) > o,p'-DDT (49.43%)

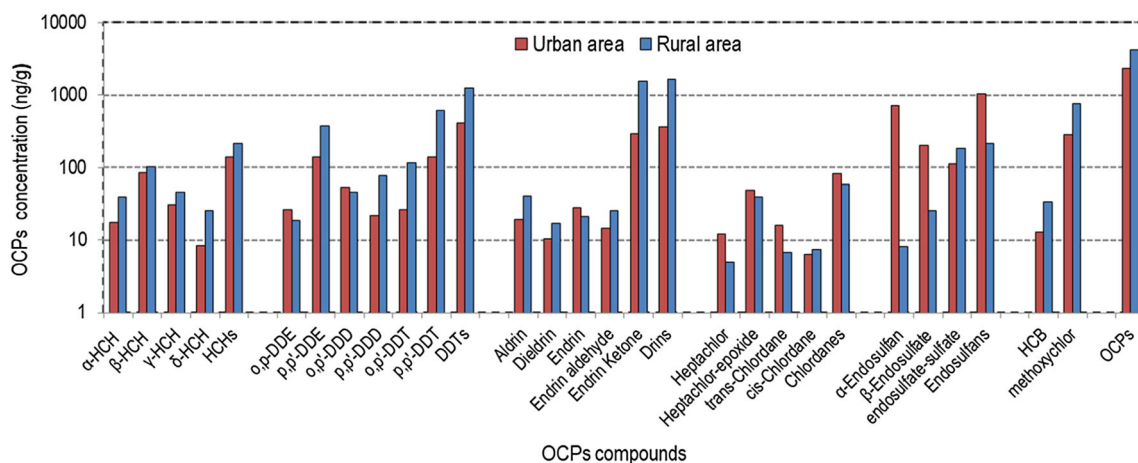


Fig. 2 Variation of individual and total (HCHs, DDTs, Drins, Chlordanes, Endosulfans, OCPs) OCP concentrations in urban areas and agricultural soils. Note the logarithmic scale applied on the Y axis

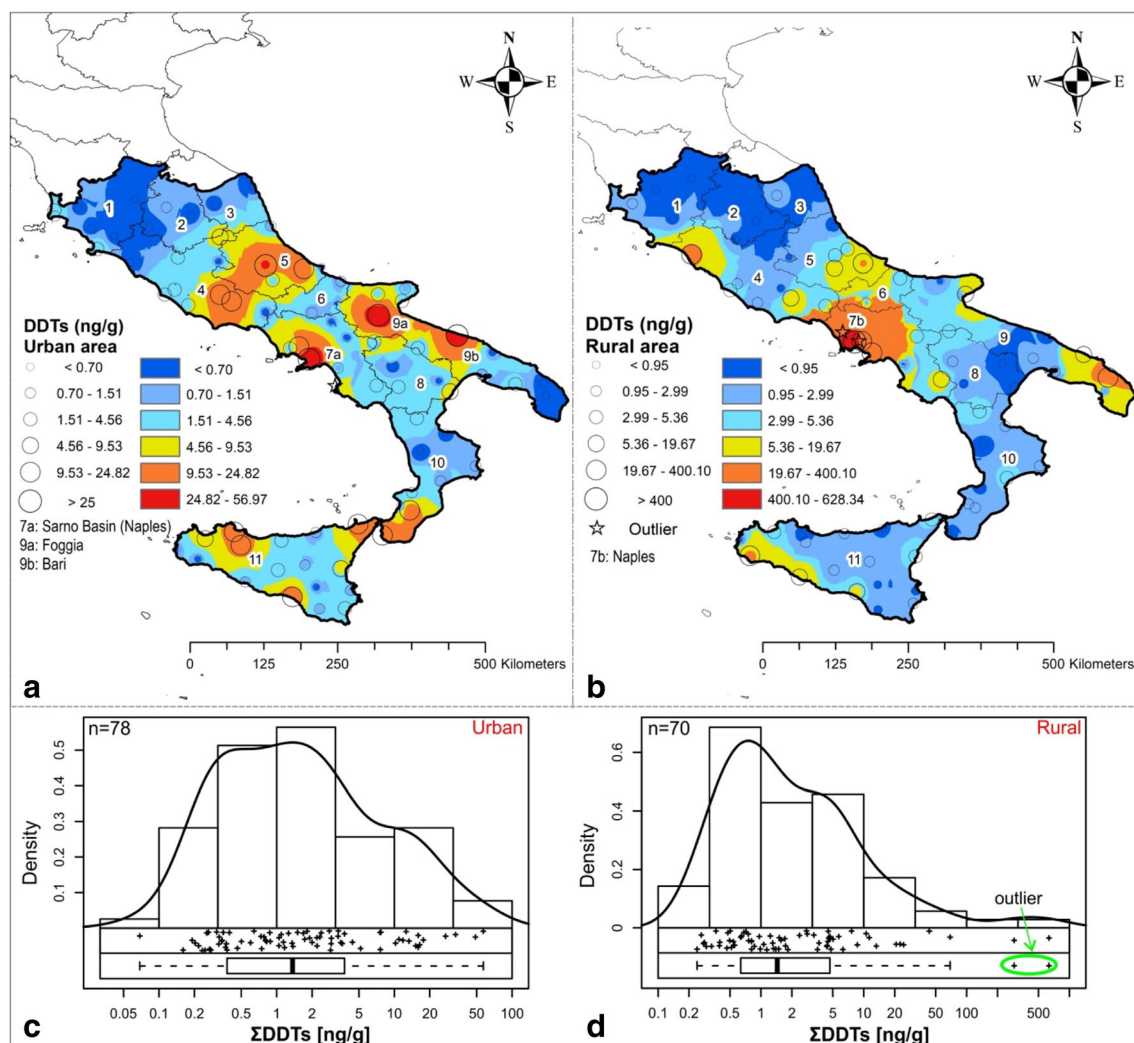


Fig. 3 Distribution of total DDTs in urban (a) and rural (b) areas. The concentration–area (C–A) fractal method was applied to set DDT ranges. Edaplots (combination of histogram, density trace, one-dimensional scattergram and Box plot) of DDTs raw data in urban (c) and rural areas (d) are displayed

When using the o,p' -DDT/ p,p' -DDT ratio (Fig. 4a), this survey highlighted a broad range of values, from 0.0002 to 214 (mean = 3.46—urban) and from 0.008 to 16.06 (mean = 0.74—rural). In general, the vast majority (92.51%) of the urban and rural sampling sites displayed a o,p' -DDT/ p,p' -DDT ratio below 7. However, high o,p' -DDT/ p,p' -DDT ratio (above 7.0) was found in some locations, mainly within urban areas. Therefore, results point towards a predominance of historical application of technical DDT with the exception of some potential recent use of dicofol for the above-mentioned highlighted urban areas.

Using the assumption that all p,p' -DDE and p,p' -DDD are degraded products of p,p' -DDT metabolite, the ratio of p,p' -DDT/(p,p' -DDE + p,p' -DDD) can be used to discern between historic applications of technical DDT (ratio < 1), compared to fresh or more recent applications (with ratio > 1) (Jiang et al. 2009). Results for this diagnostic ratio are again showing a significant range (Fig. 4b), from 0.0014 to 55.02 (mean = 4.02—urban) and from 0.006 to 40.42 (mean = 2.55—rural).

In this case, less than half of the sites (47.2%) presented a ratio below 1. When using a value of 10 as arbitrary threshold for this ratio, a large number of urban areas resulted above it. It can be derived that residues of DDT for this study can be linked to a mixed contribution from historical and recent (illegal) application. The latter was mostly highlighted in urban areas, similarly to the findings of Estellano et al. (2012), which emphasised the possible use of illegal technical DDT or dicofol in urban areas of the Tuscany region.

Total HCHs and Metabolites

Total HCH concentrations (sum of α -HCH, β -HCH, γ -HCH and δ -HCH) ranged from nd to 25.08 ng/g (mean = 1.82 ng/g—urban) and from nd to 47.27 ng/g (mean = 3.04 ng/g—rural) (Table 2). The highest values of HCHs (18.67 to 25.07 ng/g) were found in the urban areas of Bari (Apulia) (with γ -HCH isomer = 14.18 ng/g) and in the agricultural areas in the Frosinone (Latium) and Lecce (Apulia) (from 23.69 to

Table 3 Total OCP concentrations (ng/g dry weight) in topsoil of the survey area compared to those found in other studies in the recent literature

Locations	Characteristic	DDTs	HCHs	Drins	Endosulfans	Chlordanes	References
Southern Italy	Urban soils	nd–56.97	nd–25.08	nd–82.58	nd–904.2	nd–12.47	This study
Southern Italy	Rural soils	nd–632.95	nd–47.27	nd–1214.41	nd–93.13	nd–14.69	This study
Northern France	Natural areas	nd–28.6	nd–5.06	nd–2.26	nd–1.84	–	Villanneau et al. (2011)
Central Germany	Agriculture fields	23.7–173	4.6–11.5	–	–	–	Manz et al. (2001)
Southern of Poland	Urban and rural soils	23–260	1.1–11	–	–	–	Falandysz et al. (2001)
Southern of USA	Farm lands	0.10–1490	0.1–0.71	–	–	0.05–5.1	Bidleman and Leone (2004)
Zhangzhou, China	Agriculture soils	0.64–78.07	0.72–30.16	–	–	–	Yang et al. (2012)
Beijing (China)	Urban soils park	0.942–1039	0.25–197.0	–	–	–	Li et al. (2008)
Nagaon district (India)	Agriculture soils	166–2288	98–1945	–	–	–	Mishra et al. (2012)

47.11 ng/g, with β -HCH the predominant metabolite—20.37 ng/g (Fig. 5a,b). Low HCH concentrations (from nd to 2.49) were found in several areas in Tuscany, Umbria and Marche as well as in Calabria and Sicily, whilst higher values (from 2.49 to 25.07) were found in Latium, Campania and Apulia. These HCH spatial variations were well-captured by the bimodal distributions (Fig. 4c,d) indicating the existence of two different inputs or processes controlling the patterns of HCH in the study area. No outliers were recorded, but significant departure from the mean was instead highlighted (Fig. 5c,d).

The β -HCH accounted for 60.25 and 48.31% of the total HCHs, ranging from nd to 5.49 ng/g (urban) and from nd to 20.37 ng/g (rural). These values are followed by γ -HCH (21.60%) > α -HCH (12.24%) > δ -HCH (5.91%) in urban soils and γ -HCH (21.29%) > α -HCH (18.62%) > δ -HCH (11.78%) in rural soils. The dominance of β -HCH among HCH isomers may be related to its resistance to degradation and its persistence for several years in soils (Mackay et al. 1992; Calvelo Pereira et al. 2006). High residue of β -HCH isomer in rural soils of the Frosinone district (Latium) could be linked to the high contamination level of β -HCH found in the sediments of the Sacco River valley (Latium), polluted by nearby industrial landfill percolations containing by-products of Lindane (Bianconi et al. 2010; Battisti et al. 2013).

When compared to HCH concentrations in European soils such as those found in natural areas from northern France

(Villanneau et al. 2011), in agricultural soils from central Germany (Manz et al. 2001) and rural soils from southern Poland (Falandysz et al. 2001), the findings of this study reveal higher levels in comparison. On the other hand, this study presents lower levels compared to other studies, such as those related to agricultural soils of the Nagaon District (Mishra et al. 2012) and urban park of Beijing (Li et al. 2008), which highlighted HCH concentrations ranging from 98 to 1945 ng/g and 0.25 to 197 ng/g, respectively.

Technical HCHs (60–70% α -HCH, 5–12% β -HCH, 10–12% γ -HCH, 6–10% δ -HCH and impurities) and Lindane (99% γ -HCH) are two commercial pesticide compounds that are restricted for application in Italy through the European Directive in 2000 (Persistent organic pollutants amending Directive 79/117/EEC). HCH isomers have different fate and behaviour in environment. In particular, α - and γ -HCH isomers can be transformed by sunlight and through biodegradation into β -HCH, which is easily absorbed and more difficult to be evaporated from soil (Mackay et al. 1992; Calvelo Pereira et al. 2006). Studies revealed that the spatial arrangement of chlorine atoms in the β -HCH molecule protects the compound from a microbial degradation (e.g., Walker et al. 1999). To distinguish application of technical HCH from the use of Lindane, the diagnostic ratio of α/γ -HCH has been successfully used (Zhang et al. 2004), with ratios from 4.64 to 5.83 being related to application of technical HCH and

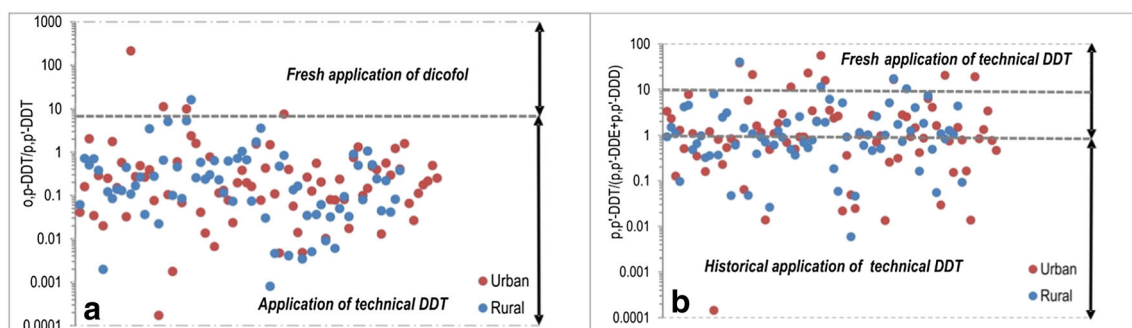


Fig. 4 Scatter diagrams of o,p'-DDT/p,p'-DDT (a) and p,p'-DDT/(p,p'-DDE + p,p'-DDD) ratios (b). The different symbology reflects whether the sites were urban (red dots) and rural nature (blue dots)

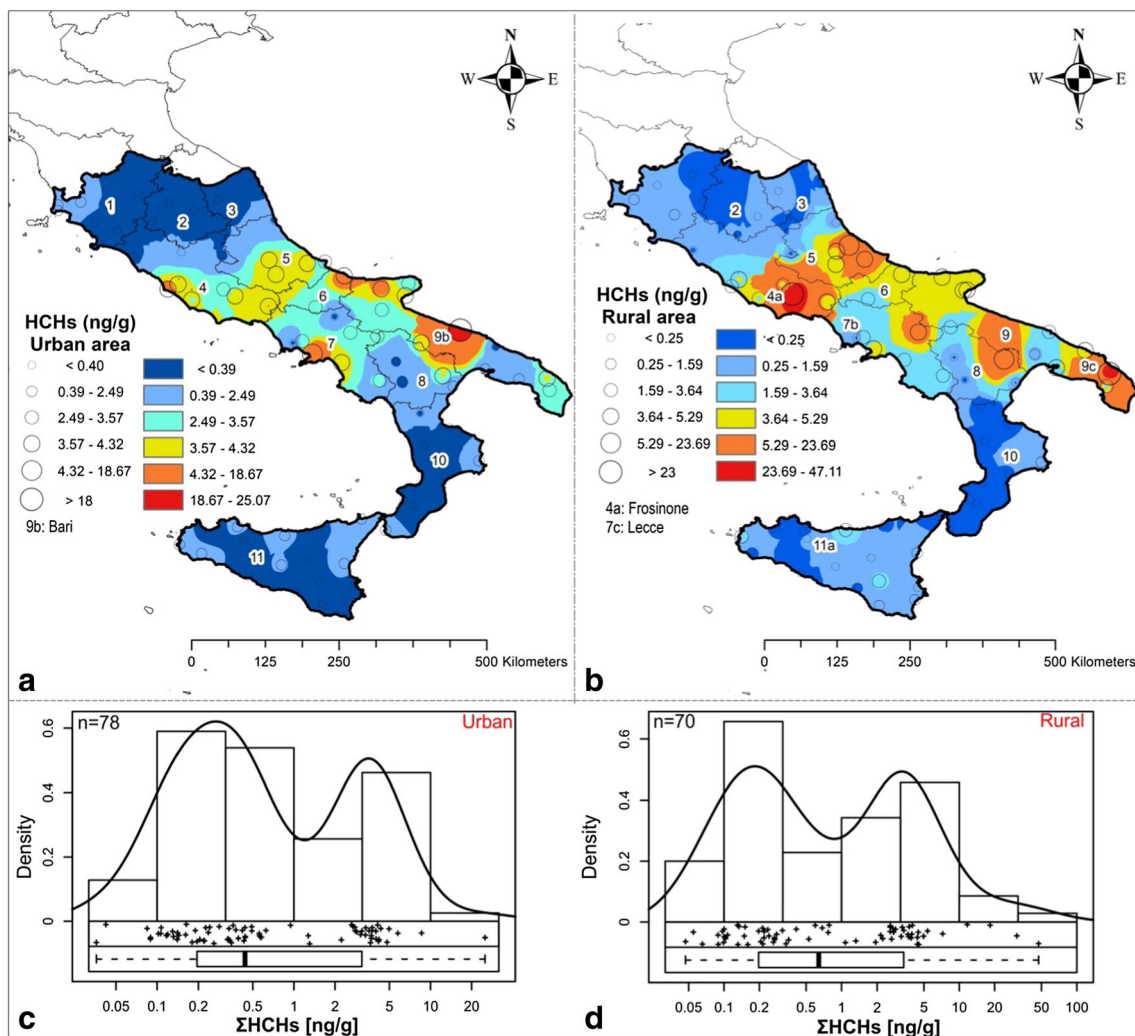


Fig. 5 Spatial distribution of HCHs in urban (**a**) and rural (**b**) areas. C–A fractal method was applied to set HCH ranges. Edaplots allow distinguishing occurrence of “outliers” observations in the survey (**c**, **d**) areas

nearly zero for Lindane applications (Zhang et al. 2004). Results for this study highlighted α/γ -HCH ratios ranging from 0.06 to 568 (mean = 12.96—urban) and from 0.09 to 78.19 (mean = 4.19—rural) (Fig. 6a). A proportion of 35.2% of the sample sites presented α/γ -HCH ratio below 1, 32.9% between 1 and 4.64, 12.2% between 4.64 and 5.83 and 9.2% a ratio above 5.83, mostly in urban areas. The 22% of the sampling sites showing a ratio of α/γ -HCH above 4.64 can possibly be linked to applications of technical DDT.

The ratios of α/β -HCH ranged from 0.002 to 822 (mean = 19.3—urban) and from 0.005 to 180 (mean = 8.21—rural) (Fig. 6b). Here, a proportion of 52.6% of the sampling sites presented α/β -HCH ratio below 1.0. The findings seem to indicate both historical application and (illegal) recent use of technical HCH in soils of the survey area. Assessment of OCPs in air samples from the Tuscany region (Estellano et al. 2012) revealed possible illegal use of technical HCH or Lindane in some urban areas.

Drins

Dieldrin, Aldrin and Endrins are collectively called Drins or Drin pesticides and were synthesized from pentadiens obtained as secondary products of petro-chemistry through the Diels-Alder reaction (Oppolzer 1991). They were primarily used as an insecticide, as well as a rodenticide and piscicide. Total Drins (sum of Dieldrin, Aldrin, Endrin, Endrin Aldehyde and Endrin Ketone) for this study ranged from nd to 82.5 ng/g (urban) and from nd to 1212 ng/g (rural). The highest urban concentrations, ranging from 31.85 to 82.5 ng/g, were found in Apulia (Bari and Foggia) and Abruzzo, whereas rural areas in the Sarno Basin (Campania) and Lecce (Apulia) presented high Drins values, ranging from 120.2 to 1212 ng/g (Fig. 7a,b). Statistically abnormal distributions and outliers were observed both in urban areas and in rural areas (Fig. 7c,d).

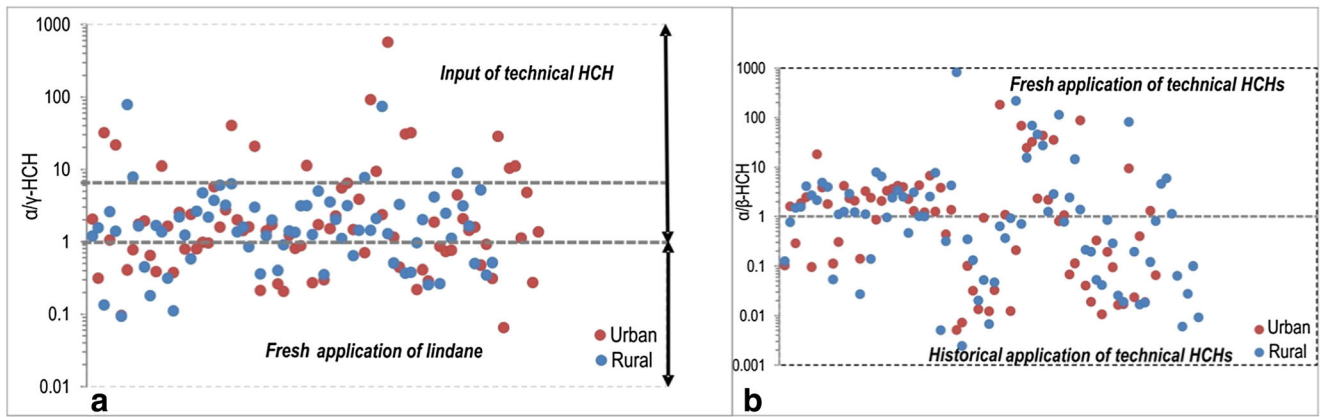


Fig. 6 Scatter diagrams of α/γ -HCH (a) and α/β -HCH (b) ratios

Among Drins, Endrin Ketone was the predominant compound accounting for 80.28% (urban) and 93.71% (rural), ranging from nd to 82.16 ng/g (urban) and from nd to 1199 ng/g (rural) (Table 2). Endrin Ketone is the final

photodegradation product of Endrin and Endrin Aldehyde and is difficult to further degrade (Fan and Alexeeff 1999). These results may indicate that the Drin residues in soils are mainly the result of historical application across the study

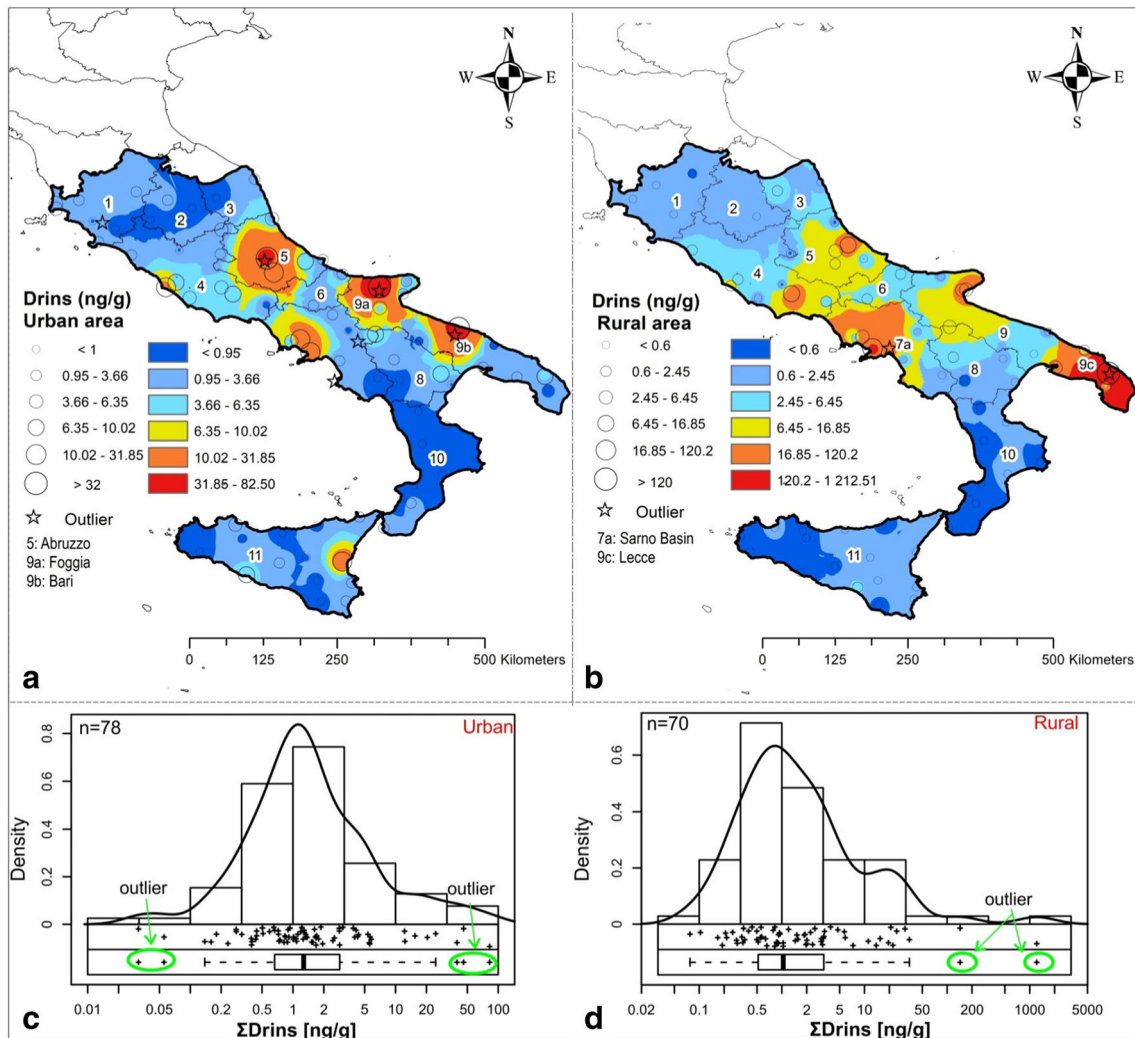


Fig. 7 Spatial distribution of Drins in urban (a) and rural (b) areas; the concentration–area (C–A) fractal method was applied to set concentration ranges. Edaplots (c, d) allow distinguishing occurrence of outlier observations in urban and rural areas

area. In comparison with other studies, for example with reported values from northern France (Villanneau et al. 2011), the present survey showed higher concentrations of Drins in urban and rural areas.

Chlordane-related compounds

Technically, Chlordane is generally used for insecticides, herbicide and termiticides and is a mixture of more than 140 related compounds (Dearth and Hites 1991). Sixty to 85% of technical chlordane is made up by stereoisomers *cis*- and *trans*-chlordane with a mixture of minor compounds such as Heptachlor, Heptachlor epoxide and *cis* and *trans*-nonachlor (Parlar et al. 1979). In this study, total concentrations of Chlordane-related compounds (sum of *cis*-chlordane, *trans*-chlordane, Heptachlor and Heptachlor-epoxide) ranged from nd to 12.46 ng/g (mean = 1.05 ng/g—urban) and from nd to 14.68 ng/g (mean = 0.84 ng/g—rural). High urban concentrations of Chlordanes were found in Campania and Bari

(Apulia), Palermo (Sicily) and Grosseto (Tuscany), ranging from 10.03 to 12.46 ng/g, whilst large rural Chlordane values were found in Tuscany, Campania (Naples) and Sicily, ranging from 6.11 to 14.68 ng/g (Fig. 8a,b). These results were confirmed by the presence of statistically abnormal distributions of Chlordane-related compounds and by one outlier (anomaly—Fig. 8c,d).

Among Chlordane-related compounds, Heptachlor epoxide was the prevalent with 58.37% (urban) and 67.56% (rural). Heptachlor epoxide is explained as an oxidation and biodegradation product of Heptachlor which has been used in the past for killing insects in households, buildings and on food crops, especially corn (Purnomo et al. 2014). Chlordane-related compounds have been banned in 1988 (ATSDR 2017). Thus, large Heptachlor epoxide concentration and mean values of Heptachlor/Heptachlor epoxide ratio equal to 0.23 (urban) and 0.14 (rural) point towards historical application of the commercial Chlordane. However, when compared to similar studies, such as that conducted by Bidleman and

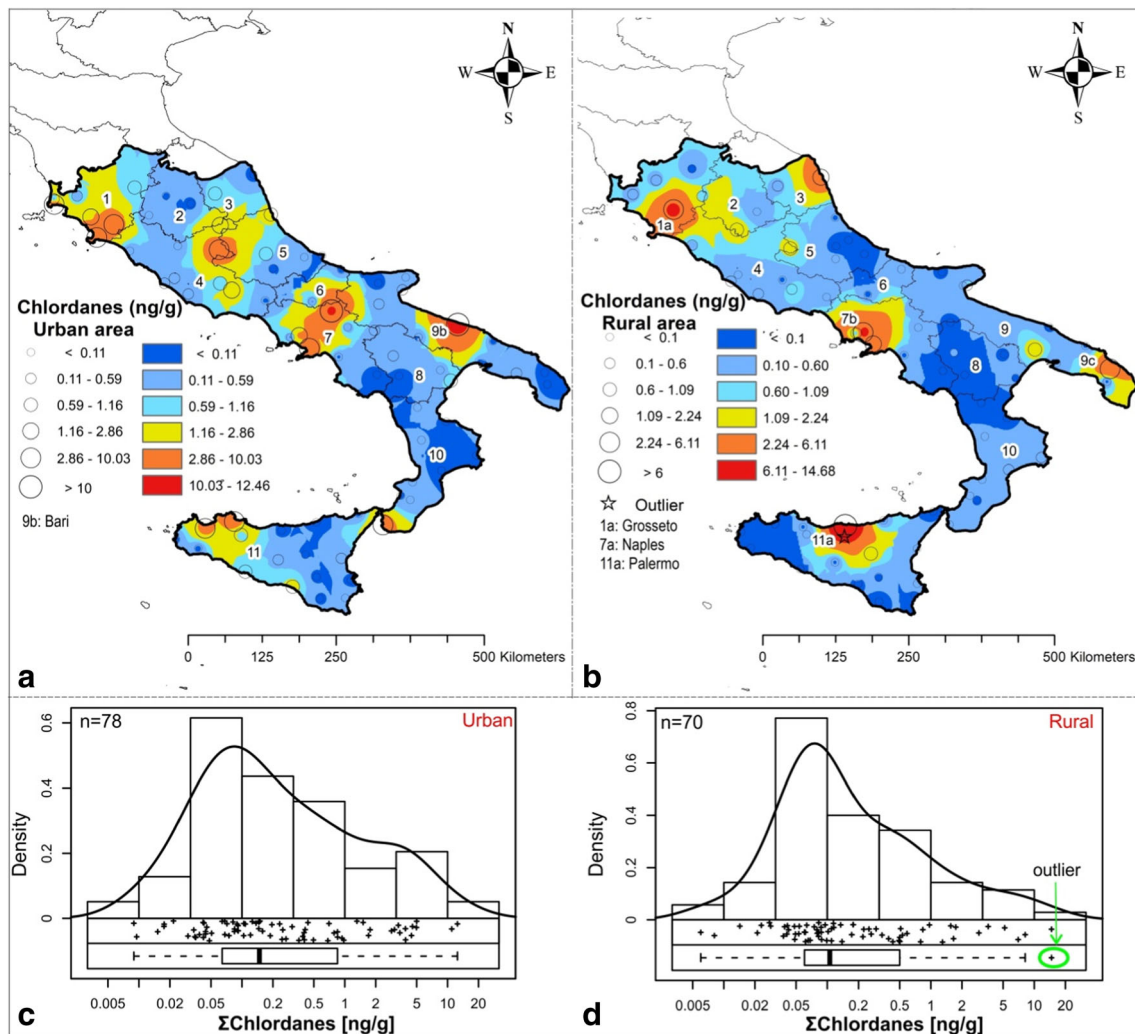


Fig. 8 Distribution of total Chlordane-related compounds in urban (a) and rural (b) areas. Edaplots (c,d) reveal abnormal distribution of the dataset and occurrence of an outlier rural area

Leone (2004) in farmland of the Southern of USA (Chlordane-related compound concentration ranging from 0.05 to 5.1 ng/g), the results from this study seem to suggest extremely extensive applications made in some parts of the studied area.

Endosulfans

Endosulfan is a cyclodiene pesticide used worldwide to control pests in non-food crops (cotton, tobacco, timber and ornamental plants) and food crops such as vegetables, fruits, corn and cereals and to control a wide variety of insects and mites (ATSDR 2000). Italy is the second consumer of Endosulfan in European Union with 20% of the total volume, after Spain (Endosulfan Preliminary Dossier 2003). Technical endosulfan was globally banned under the Stockholm Convention (2011) because of its threats to human health and the environment. Endosulfan is made up of α - and β -endosulfan isomers that are fairly resistant to degradation and persistent in the environment. Endosulfan sulfate is the degradation product of Endosulfan, and it is a more hydro-soluble metabolite and susceptible to photolysis (Cerrillo et al. 2005).

In this survey, total Endosulfan (sum of α -endosulfan, β -endosulfan and endosulfan sulfate) ranged from nd to 904.21 ng/g (mean = 13.25 ng/g) accounting for 44.32% of the total OCPs in urban areas and from nd to 92.99 ng/g (mean = 3.08 ng/g) accounting for 5.12% of total OCPs in rural areas. High Endosulfan concentrations were found in the urban area of Bari (Apulia), ranging from 71 to 904.21 ng/g and in rural areas of Lecce (Apulia) from 55.32 to 92.99 ng/g (Fig. 9a,b). These values are extremely large if compared to those found in natural areas of the Northern France (ranging from nd to 1.84 ng/g; Villanneau et al. 2011). Statistical distributions showed both outliers as well as abnormal behaviour of Endosulfan concentrations (Fig. 9c,d), which could be associated with the diverse chemical processes that may affect endosulfan compound behaviour in soils medium. Since α -endosulfan decomposes more easily than β -endosulfan in soil, the ratio of α/β -endosulfan < 2.33 may be used to judge the age of their residues in soil (Jennings and Li 2014; Jia et al. 2010). In urban areas, α -endosulfan isomer constituted 69.59% of the total endosulfan followed by β -endosulfan with 19.36% and endosulfan sulphate (11.05%), and the ratio of α/β -endosulfan ranged from 0.05 to 312.9 (mean = 22.44). Endosulfan sulphate was the predominant compound in rural areas (84.58%), followed by β -endosulfan and α -endosulfan, and the ratio of α/β -endosulfan ranged from nd to 40 (mean = 1.59). These results strongly suggest a recent (illegal) use of technical endosulfan in urban areas, especially in Apulia. In contrast, results for rural areas seem to point to historical application. The relatively recent restriction of technical endosulfan (Stockholm

Convention 2011) and its uses in Italy until December 2007 may explain why it was still found in high proportion in the soils of the survey area (Pozo et al. 2016; Qu et al. 2017).

HCB and Methoxychlor

HCB was listed among the first group of persistent OCP compounds in the Stockholm Convention (2005), even though it has been restricted since 1985 in the European Union countries (Barber et al. 2005). It has been used as fungicide to control bunt on wheat and seed treatment of onions and sorghum (Courtney 1979). The values of HCB in this survey ranged from 0.01 to 2.39 ng/g (mean = 0.16 ng/g—urban) and from nd to 13.37 ng/g (mean = 0.47 ng/g—rural). HCB made up 0.55% (urban) and 0.79% (rural) of the total OCP concentrations. Several studies reported that HCB is still used as a by-product or impurity in several chemical compounds, including chlorinated pesticides, such as Lindane (Pacyna et al. 2003; Barber et al. 2005). Pearson correlation coefficient between HCB and γ -HCH compounds showed a slight correlation ($r = 0.44$), which may suggest that HCB could be partially related to input of technical HCH or Lindane in the study area.

Most Methoxychlor enters the environment when it is applied to forests, agricultural crops and farm animals as insecticide (US EPA 1991). It is one of the few organochlorine pesticides that has undergone an increase in its use since the ban on DDT, but Methoxychlor was finally listed as banned OCP pesticides by the United Nations Environmental Program (UNEP) (Stockholm Convention 2011). In this study, the concentrations of Methoxychlor ranged from nd to 53 ng/g (mean = 3.64 ng/g—urban) and from nd to 521 ng/g (mean of 10.96 ng/g—rural). When compared to other studies, the mean concentration of Methoxychlor (10.96) found in rural areas is comparable to that from agricultural soils of central China (Zhou et al. 2013), but bigger than that found in southern Mexico (Cantu-Soto et al. 2011) and soils from the hilly areas of Nepal (Yadav et al. 2017).

Compositional biplot and robust factor analysis

Compositional biplots explained 66.9% (PC1-PC2) and 61.5% (PC1-PC3) of the variability (Fig. 10).

The sum of DDTs (DDTs) p,p'-DDT, o,p'-DDT, p,p'-DDE, o,p'-DDE, p,p'-DDD and o,p'-DDD may be considered a variable association (A), due to the vicinity of their vertices and their rays pointing to the same direction. Among the associations, DDTs and p,p'-DDE displayed the highest vector length (communality—Fig. 10a). It was possible to discriminate DDT variables and highlight two sub-groups based on their chemical structures (Fig. 10b):

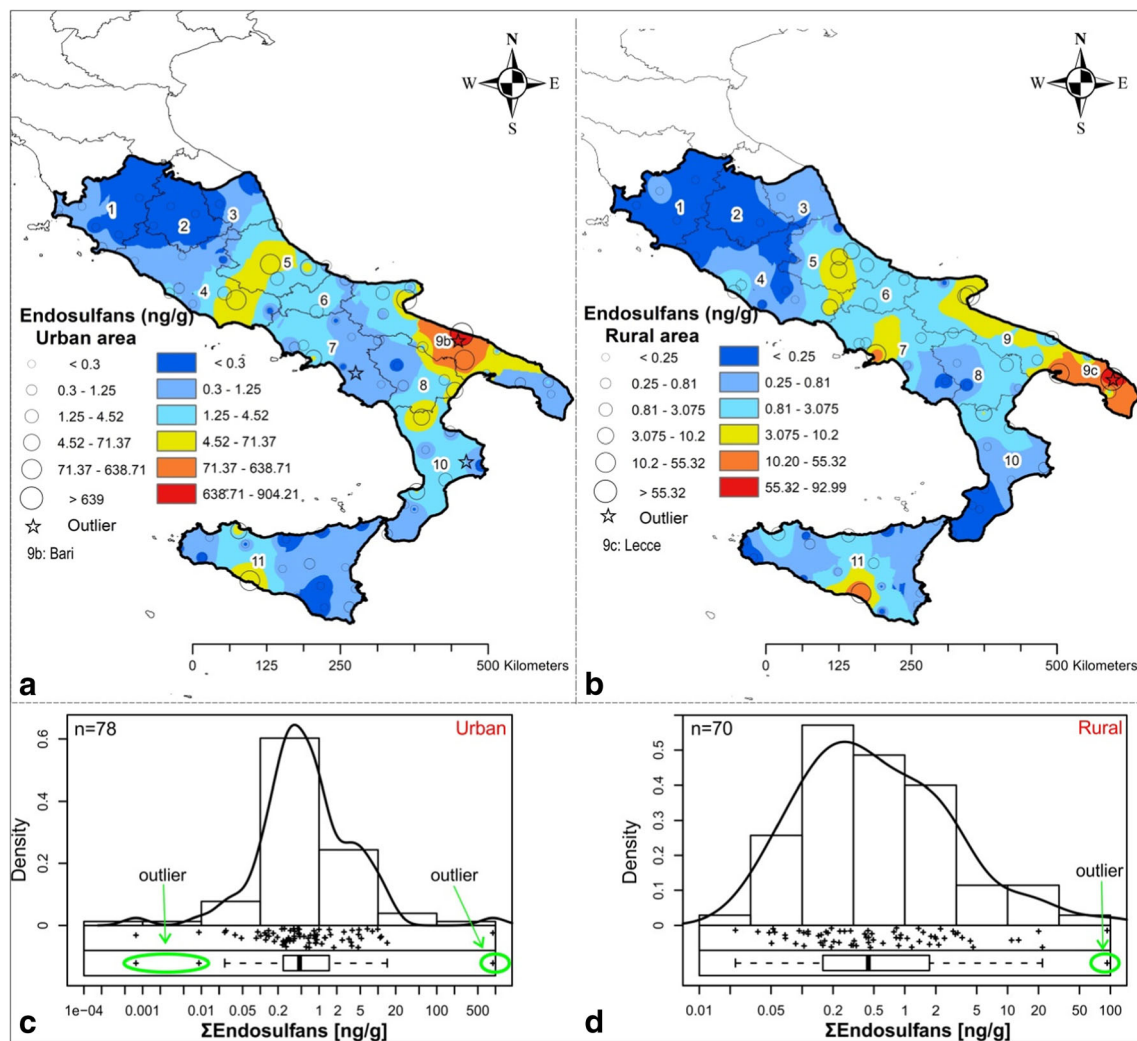


Fig. 9 Distribution of concentration of Endosulfan in urban (a) and rural (b) areas. Edaplots (c,d) reveal abnormal distribution of the dataset and occurrence of low and high outliers in urban and rural areas

- A_1 , formed by o,p'-DDT and o,p'-DDD (superimposition of vertices and low communalities, and
- A_2 , formed by p,p'-DDT, p,p'-DDE and p,p'-DDD (proximity and high communality).

Furthermore, the o,p'-DDE variable presented a high length vector and it was separated from other DDT metabolites. These findings point towards a significant interrelationship between DDT isomers. High lengths of p,p'-DDT metabolites could be associated with a dominant input of technical DDT in soils of the study area (Qu et al. 2016). Low communalities and superimposition of the o,p'-DDT and o,p'-DDD variables may also be associated with the use of dicofol which contains more o,p'-DDT metabolites (Qiu et al. 2005). Moreover, the high length of the o,p'-DDE vector and its disassociation to other DDT isomers may illustrate its specific behaviour in soil based on its specific physico-chemical properties (solubility, partition coefficient and vapour pressure) (Pfaender and Alexander 1972).

Another variable association (B) made by α -HCH, γ -HCH and δ -HCH was highlighted (Fig. 10a). This was also characterized by the vicinity of their vertices with their vectors being superimposed to one another. This configuration can be associated to a similar behaviour of these three isomers in the soils of the study area. In fact, α -HCH, γ -HCH and δ -HCH metabolites are the main compounds related to the commercial technical HCH (Senthil Kumar et al. 2001), which is reflected by how α -HCH and δ -HCH vectors are geometrically symmetric with respect to the HCH vector (Fig. 10b). This spatial configuration illustrates their similar behaviour in the soils of the study area. The β -HCH variable is marked by a high communality, and its vector forms 90° with the (B) variable association (Fig. 10a). These results seem to suggest that the β -HCH compound has a different fate and behaviour in soils compared to other HCH metabolites. Moreover, the high communality and disassociation of β -HCH with respect to other HCH metabolites could explain its persistence to degradation and its accumulation in soil of the study area (Jiang et al. 2009).

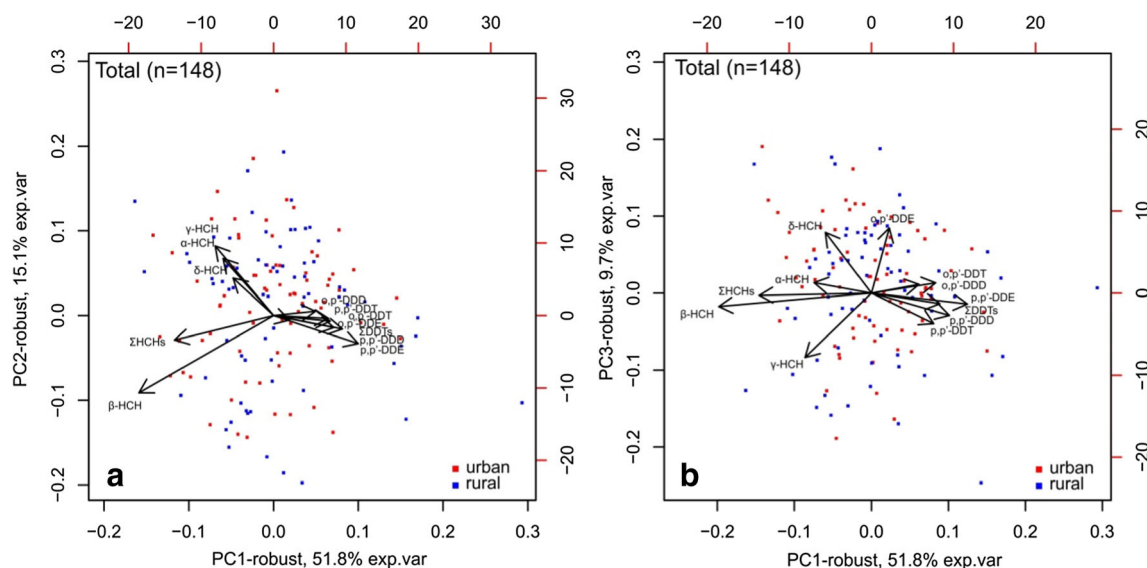


Fig. 10 Robust biplots for the first and second principal components (a) and for the first and third principal components (b) based on DDTs and HCHs investigated compounds

Factor analysis was performed to determine the correlation between DDT and HCH isomers which further revealed the possible sources of these compounds. Factor loadings and total variances of individual OCPs were computed to facilitate the interpretation (Table 4). DDT and HCH metabolites of the three-factor model were separated by positive and negative loadings and sorted in descending order:

- Urban sites

F₁: o,p'-DDT, p,p'-DDD, -(γ-HCH, α-HCH)

F₂: p,p'-DDD-(δ-HCH, β-HCH)

F₃: p,p'-DDT, -(o,p'-DDE)

- Rural sites

F₁: p,p'-DDD, o,p'-DDT, o,p'-DDD, -(γ-HCH, α-HCH)

F₂: p,p'-DDE, o,p'-DDE, -(β-HCH)

F₃: o,p'-DDD, -(p,p'-DDT)

Factor score values for F₁, ranging from −3.78 to 1.64 (urban) and −2.75 to 2.94 (rural), were plotted to represent their spatial distribution (Fig. 11a, b). High urban factor score values (ranging from 1.26 to 1.62), associated mainly with o,p'-DDT (0.84) and p,p'-DDD (0.76) compounds, were found in Frosinone (Latium), Foggia (Apulia), southeastern coastal area of Calabria and the Sicily region (Palermo and Gela) (Fig. 11a). High-factor loading of the o,p'-DDT isomer (0.84), explained by its dominance in urban soils, may be attributed to the application of dicofol, containing high o,p'-DDT residue. Further increase of p,p'-DDD (0.76) isomer in these areas was relatively significant and might be related to

degradation processes of DDT compounds. Low urban factor scores (< to −1.60), tied to α-HCH (−0.69) and γ-HCH (−0.68), were mainly observed in Calabria and Marches. These are potentially related to application of technical HCH. The physico-chemical properties of α-HCH and γ-HCH are similar, showing a relatively easy degradation in soils (Mackay et al. 1992; Calvelo Pereira et al. 2006).

The highest rural factor score values (ranging from 1.52 to 2.94), associated with p,p'-DDD (0.81), o,p'-DDT (0.79) and

Table 4 Varimax-rotated factor (three-factor model) using 78 topsoil samples from urban areas and 70 samples from agricultural soils; bold entries: loading values over |0.50|

Variables	Urban areas			Rural areas		
	F1	F2	F3	F1	F2	F3
α-HCH	−0.69	0.12	−0.36	−0.80	−0.08	0.19
β-HCH	0.10	−0.83	−0.12	−0.15	−0.77	−0.06
γ-HCH	−0.68	0.04	−0.19	−0.71	−0.26	0.03
δ-HCH	−0.25	−0.64	0.17	−0.24	−0.43	0.31
o,p'-DDE	−0.33	0.04	−0.75	−0.45	0.53	0.33
p,p'-DDE	0.04	0.81	0.12	0.16	0.81	−0.30
o,p'-DDD	0.45	0.33	0.47	0.64	0.15	0.62
p,p'-DDD	0.76	0.19	0.03	0.81	0.28	−0.10
o,p'-DDT	0.84	0.13	−0.13	0.79	0.03	−0.09
p,p'-DDT	−0.08	0.07	0.79	0.30	0.25	−0.81
Eigenvalues	2.60	1.96	1.64	3.77	1.675	1.23
Total variance in %	26.04	19.58	16.45	33.31	19.6	13.81
Cum. of total variance	26.04	45.63	62.09	33.31	52.92	66.81

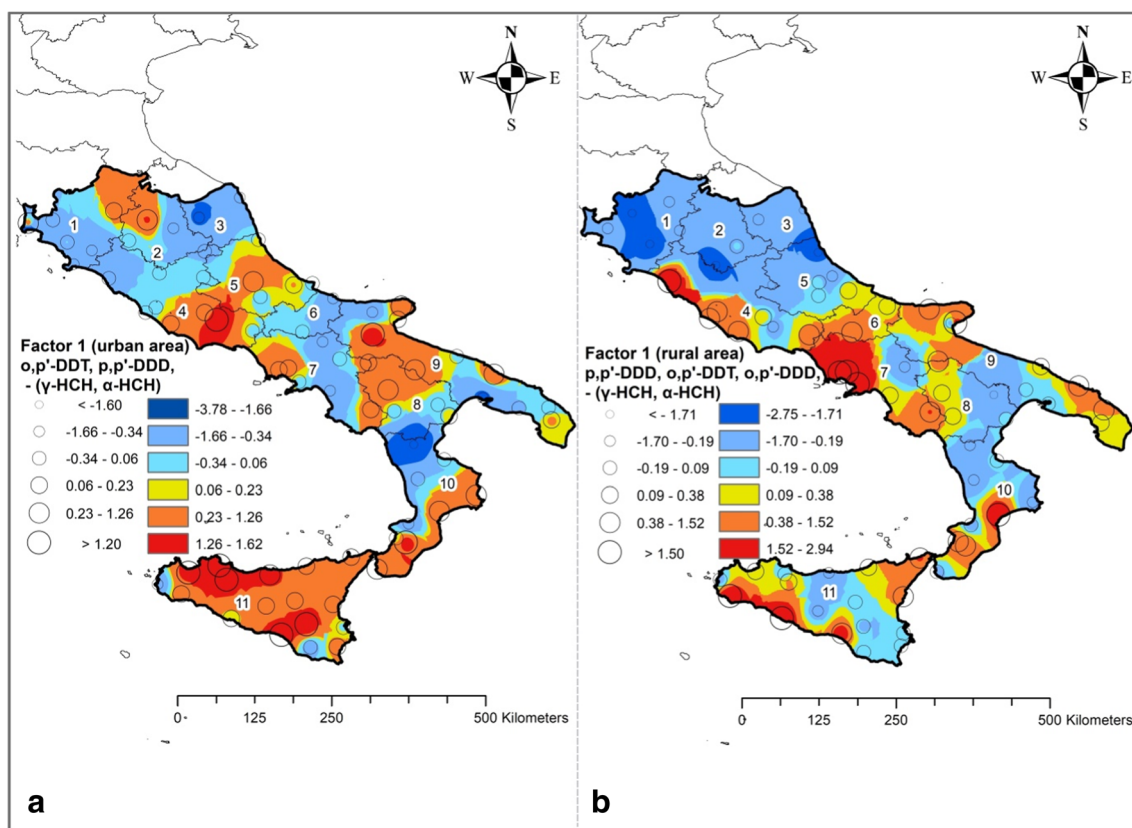


Fig. 11 Dots and interpolated factor score maps of factor 1 in urban areas (**a**) and rural soils (**b**). Factor score value ranges are based on C–A (concentration–area) fractal plot after a min–max normalisation

o,p'-DDD (0.64) compounds, were found mainly along the coasts (Latium, Campania, Naples, Calabria and southern Sicily—Fig. 11b), where intensive agriculture activities such as those carried out in vineyards and olive plantations occur (Corona et al. 2012; ISPRA 2014a). The higher loading of p,p'-DDD compound can be associated to historical applications of DDT together with a more recent application of dicofol, illustrated by occurrence of o,p'-DDT and o,p'-DDD isomers in these areas. This is partially in line with the results of Qu et al. (2016) which have indicated that DDT residues in the Campania plain are mainly the result of historical application.

F2 factor score values ranged from −2.56 to 2.26 (urban) and −2.16 to 1.92 (rural) (Fig. 12a, b). The urban areas of Grosseto (Tuscany), northern Campania and Taranto (Apulia) displayed the highest factor scores (> 1.90) corresponding to the p,p'-DDE (0.81) isomer (Fig. 12a). These results can be associated to historical application of technical DDT because p,p'-DDE is a degradation product of p,p'-DDT isomer.

The highest F₂ factor scores (ranging from 1.77 to 2.75) were related to p,p'-DDE (0.81) and o,p'-DDE (0.53) and were found in most rural sites in the northern

part of the study area and in Naples (Campania). This may be attributed to historical application of technical DDT, because high-factor scores of DDE metabolites are matching their degradations and fate in situ. Low-factor score values (ranging from −2.16 to −1.39) corresponded to β-HCH (−0.77) and were found in rural sites of Frosinone (Latium) (Fig. 12b). This might be related to the dominance or specific behaviour of the β-HCH metabolite in soils of this area. As previously mentioned, high level of the β-HCH isomer was found in soils and sediments from the Sacco River valley (Frosinone), which are polluted by the release of industrial landfill percolations containing by-product of Lindane.

Values for the F₃ factor score, ranged from −2.83 to 2.05 (urban) and −3.93 to 2.66 (rural) (Fig. 13). The highest urban factor score values (> 1.23) corresponded to p,p'-DDT (0.79) and were found in Palermo (Sicily), Naples (Campania) and Tuscany (Fig. 13a). This is likely to be related to recent illegal application of technical DDT, which is confirmed by the dominance of the p,p'-DDT isomer in soils of these areas, similarly to what Qu et al. (2016) have found in soils of the Campania plain. Pozzo et al. (2016) also highlighted recent use of technical DDT in Palermo (Sicily), whilst an assessment of OCPs

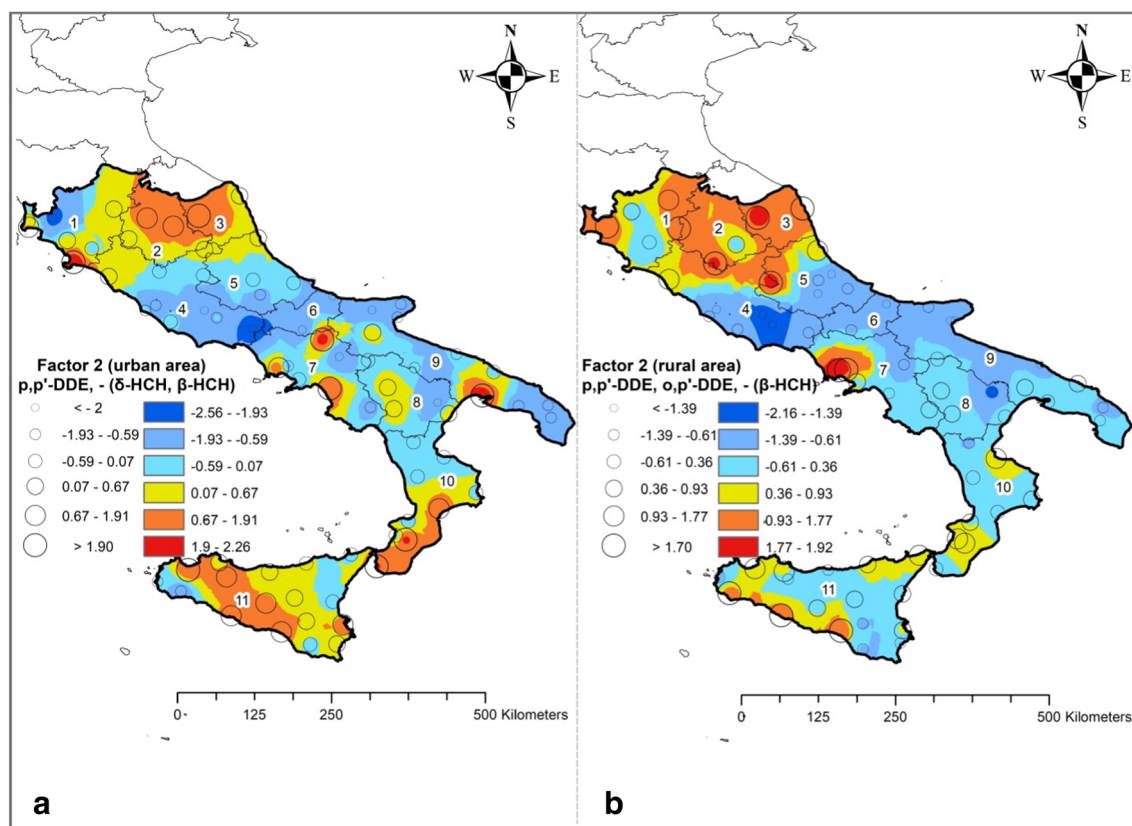


Fig. 12 Dots and interpolated factor score maps of factor 2 in urban areas (a) and rural soils (b). Factor score value ranges are based on C–A fractal plot after a min–max normalisation

pollution sources in urban air of Tuscany revealed possible illegal use of commercial technical DDT (Estellano et al. 2012). Low-factor score values (ranging from -2.83 to -1.85) were found in Tuscany and corresponded to o,p' -DDE. We might preclude that the occurrence o,p' -DDE isomer in Tuscany may be related to unknown synthetic chemicals. A follow-up study in this region may give reasons of the occurrence of this metabolite in Tuscany soils.

The highest rural factor score values (ranging from 2.09 to 2.66) corresponded to o,p' -DDD in Calabria (Cosenza) and Tuscany (Fig. 13). Low values (ranging from -3.93 to -1.67) were found in rural areas in Basilicata, revealing dominance of p,p' -DDT. These results point towards a mixed input of DDT residues through recent use and historical application in rural areas of Calabria.

Contamination assessment

The SoQI index was used to represent the degree of contamination and, therefore, concern of the studied area (Fig. 14).

SoQI values in soils of urban and rural areas of the Tuscany, Umbria, Marches and Molise are equal to 100

(Fig. 14a,b). This is associated to very low contamination levels, where none of the sampling sites (both in urban and rural soils) presented concentration beyond the threshold values established by Italian environmental legislation (D. Lgs. 152/2006). Similarly, low levels of concern were observed for urban and rural areas of Basilicata and Calabria (Fig. 14a,b).

Urban soils in Campania (Naples, Sarno Basin), Abruzzo and Apulia (Foggia and Bari) presented SoQI ranging from 50 to 70, corresponding to a medium contamination level. In addition, rural soils in Latium (Frosinone and Civitavecchia), Abruzzo and Apulia (Taranto and Manfredonia) also showed the same SoQI (50–70).

The lowest urban SoQI value (46.2), associated to a high contamination level, was found in urban soils of the metropolitan area of Foggia. This could be related to the use of OCP pesticides against pests in urban gardens. For rural areas, instead, the lowest SoQI values (ranging from 30 to 50) were found in soils of Campania (Naples metropolitan area and Sarno Basin) and Apulia (Lecce). This further confirms the observations made previously which linked high contamination levels with intensive agricultural activities such as those

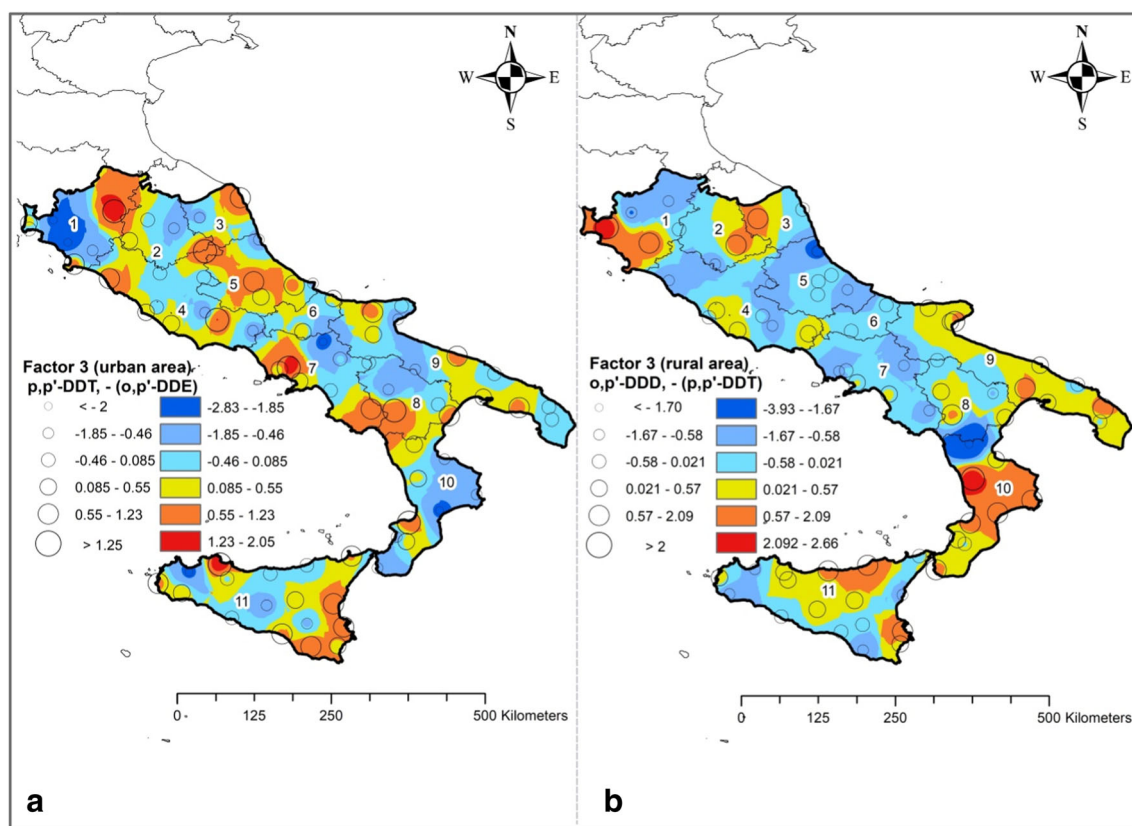


Fig. 13 Dots and interpolated factor score maps of factor 3 in urban areas (a) and rural soils (b). Factor score value ranges are based on C–A fractal plot after a min–max normalisation

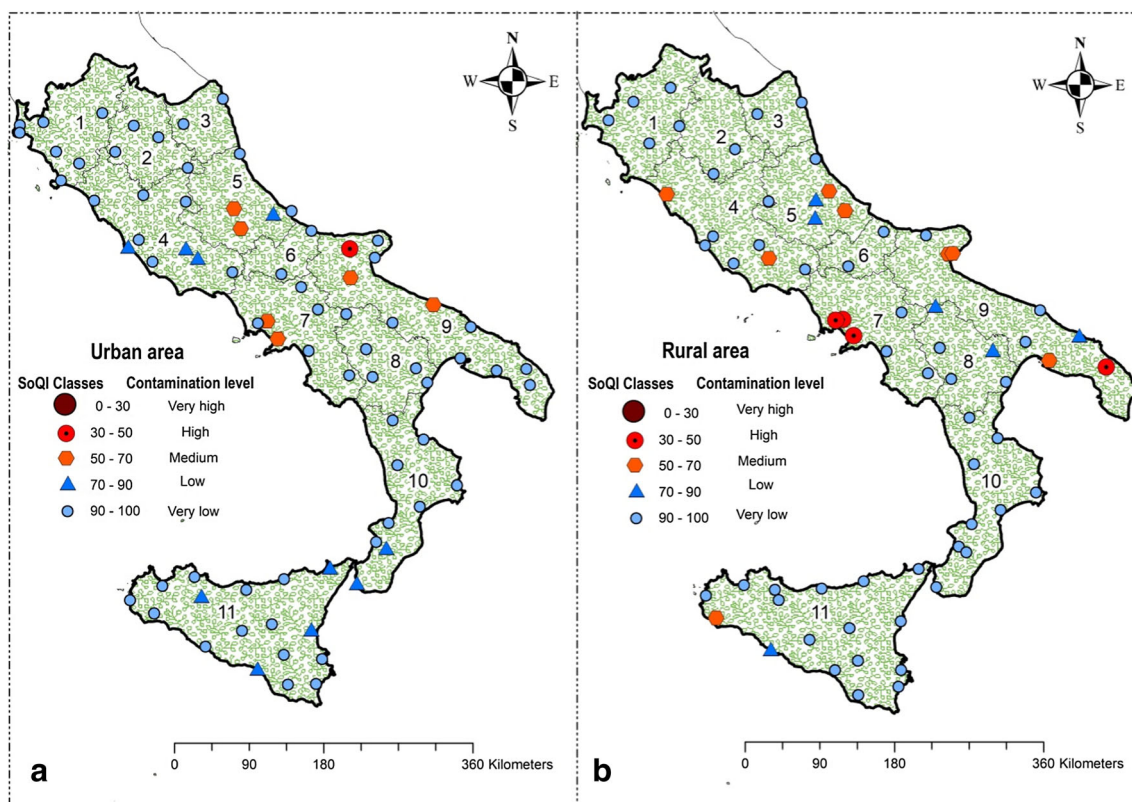


Fig. 14 Spatial distribution of SoQI contamination levels in urban (a) and rural areas (b) in the 11 regions of the study area

occurring in vineyards and olive plantation along the coastal areas.

Conclusion

This study presented the results of a regional survey of OCP compounds conducted in urban and rural soils of 11 administrative regions from central and southern Italy, as part of an ongoing project aiming to cover the entire Italian territory. The main findings revealed the concentration of 24 OCPs, ranging from nd to 1043 ng/g (mean = 29.91 ng/g—urban) and from nd to 1914 ng/g (mean of 60.16 ng/g—rural). In particular, high DDT concentrations were mostly shown in urban and rural soils of Campania and Apulia. Enrichment of HCH was also highlighted in the central regions of the study area, with relatively lower values in the northern and southern parts. Furthermore, Endosulfan-related compounds and Methoxychlor were found to be 42.32 and 12.17% of the total OCPs in urban areas, respectively, which are likely to be related to recent applications, particularly in Apulia.

Diagnostic ratios of DDT residues clearly unveiled a dominance of historical application of these compounds in soil, but also a minor (yet significant) more recent illegal use of technical DDT or dicofol mainly in urban areas. A mixed application of HCH was also highlighted, with residues both from historical and recent applications. On the other hand, the different compositions of Drins and Chlordane-related compounds emphasised that the residues of these compounds are mainly related to a historical application. At the same time, recent applications of Endosulfan residues in urban areas were suggested, together with a historical use of this compound in agricultural soils. Unfortunately, the Italian environmental legislation has not established to date any guideline (threshold) values with regard to Endosulfan residues in soil (see D. Lgs. 152/2006), failing to recognise their potential threat to human health.

These results were also backed up by the findings of the multivariate computations performed on DDT and HCH residues, pointing out that (1) DDT and HCH residues could be mainly related to historical but also more recent (illegal) application; (2) occurrence of DDT residues in soils of the Campania region could be related to historical application of technical DDT; (3) indirect evidence of illegal “fresh” application of DDT was identified in urban areas of Tuscany, Sicily and Campania; (4) HCH levels in Latium (Frosinone) rural areas could be related to β -HCH metabolite in the anomalous sediments of the Sacco River valley, affected by nearby industrial landfill percolations.

This study should be considered as a first stepping stone (as a regional survey) towards a major investigation on the main sources and levels of OCPs throughout the Italian territory. As such, it is envisaged that the findings will contribute to build

OCP baseline and drive towards an entire coverage of the Italian territory. The survey, which is currently progressing in the remaining 9 regions of northern Italy, has highlighted areas with high concentrations of some OCPs, which can be in part explained by recent (illegal) applications. Even though it was not the scope of this study, the results highlighted some potential human health concerns, which need addressing urgently. Given the associated human health risks and the potential wider implications for the environment, these results strongly point towards follow-up studies to be held in areas of higher contamination levels (Naples and Sarno Basin as well as Foggia and Lecce regions), with a larger number and higher density of soil and air samples. It is also hoped that similar studies will build science-based evidence to be fed back at institutional level for more adequate and comprehensive regulations and, in the long-term, for a full ratification of the Stockholm Convention.

Acknowledgments This work was sustained by the financial support of the MIUR through the following two industrial projects: (1) “Integrated agro-industrial chains with high energy efficiency for the development of eco-compatible processes of energy and biochemicals production from renewable sources and for the land valorisation (EnerbioChem)” PON01_01966, funded in the framework of the Operative National Program Research and Competitiveness 2007–2013 D. D. Prot. n. 01/Ric. 18.1.2010 and (2) “Development of green technologies for production of BIOchemicals and their use in preparation and industrial application of POLImeric materials from agricultural biomasses cultivated in a sustainable way in Campania Region (BioPoliS)” PON03PE_00107_1, funded in the framework of the Operative National Program Research and Competitiveness 2007–2013 D. D. Prot. N. 713/Ric. 29/10/2010 (Research Units Responsible: Prof. B. De Vivo).

References

- Albanese S, De Vivo B, Lima A, Cicchella D (2007) Geochemical background and baseline values of toxic elements in stream sediments of Campania region (Italy). *J Geochem Explor* 93(1):21–34
- Alves AAR et al (2012) Comparison between GC-MS-SIM and GC-ECD for the determination of residues of organochlorine and organophosphorus pesticides in Brazilian citrus essential oils. *J Braz Chem Soc* 23(2):306–314. <https://doi.org/10.1590/S0103-50532012000200017> ISSN 0103-5053. ATSDR (Agency for Toxic Substances and Disease Registry) 1995. Chlordane (CAS# 12789-03-6). <http://www.atsdr.cdc.gov/toxfaqs/tfacts31.pdf> (Feb. 26, 2014)
- APAT (Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici) (2008) Criteri metodologici per l'applicazione dell'analisi assoluta di rischio ai siti contaminati. Revisione 2. APAT, Roma, 156 pp. <http://www.isprambiente.gov.it/files/temi/siti-contaminati-02marzo08.pdf>
- Aramendia MA, Borau V, Lafont F, Marinas A, Marinas JM, Moreno JM, Urbano FJ (2007) Determination of herbicide residues in olive oil by gas chromatography–tandem mass spectrometry. *Food Chem* 105(2):855–861
- Arienzo M, Albanese S, Lima A, Cannatelli C, Aliberti F, Cicotti F, Qi S, De Vivo B (2015) Assessment of the concentrations of polycyclic aromatic hydrocarbons and organochlorine pesticides in soils from the Sarno River basin, Italy, and ecotoxicological survey by *Daphnia magna*. *Environ Monit Assess* 187:1–14

- ATSDR (Agency for Toxic Substances and Disease Registry, USA) (2000) Toxicological profile for endosulfan. ATSDR, Atlanta
- ATSDR (2017) Chlordane. Full SPL data. Substance priority list (SPL) resource page. Agency for Toxic Substances and Disease Registry, Centers for Disease Control and Prevention. <http://www.atsdr.cdc.gov/SPL/resources/index.html>. Accessed 6 Oct 2017
- Barber JL, Sweetman AJ, Van Wijk D, Jones KC (2005) Hexachlorobenzene in the global environment: emissions, levels, distribution, trends and processes. *Sci Total Environ* 349:1–44
- Battisti S, Caminiti A, Ciotoli G, Panetta V, Rombolà P, Sala M, Ubaldi A, Scaramozzino P (2013) A spatial, statistical approach to map the risk of milk contamination by beta-hexachlorocyclohexane in dairy farms. *Geospat Health* 8(1):77–86
- Bianconi D, De Paolis MR, Agnello MC, Lippi D, Pietrini F, Zacchini M, Polcaro C, Donati E, Paris P, Spina S, Massacci A (2010) Field-scale rhizoremediation of contaminated soil with hexachlorocyclohexane (HCH) isomers: the potential of poplars for environmental restoration and economical sustainability. In: Golubev IA (ed) *Handbook of phytoremediation*. Nova Science Publishers Inc., Hauppauge, pp 783–794
- Bidleman TF, Leone AD (2004) Soil-air exchange of organochlorine pesticides in the Southern United States (US). *Environ Pollut* 128:49–57
- Calvelo Pereira R, Camps-Arbestain M, Rodríguez Garrido B, Macías F, Monterroso C (2006) Behaviour of α -, β -, γ - and δ -hexachlorocyclohexane in the soil-plant system of a contaminated site. *Environ Pollut* 144:210–217
- Cantu-Soto EU, Meza-Montenegro MM, Valenzuela-Quintanar AI, Felix-Fuentes A, Grajeda-Cota P, Balderas-Cortes JJ et al (2011) Residues of organochlorine pesticides in soils from the Southern Sonora, Mexico. *Bull Environ Contam Toxicol* 87(5):556–560
- CCME (1992) National classification system for contaminated sites. Report # CCME EPC-CS39E. Canadian Council of Ministers of the Environment, Winnipeg
- CCME (Canadian Council of Ministers of the Environment). (2007) Soil quality index 1.0: technical report. In: Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg
- Cerrillo I, Granada A, Lopez-Espinosa MJ, Olmos B, Jimenez M, Cano A, Olea N, Olea-Serrano MF (2005) Endosulfan and its metabolites in fertile women, placenta, cord blood, and human milk. *Environ Res* 98:233–239
- Cheng Q, Agterberg FP, Ballantyne SB (1994) The separation of geochemical anomalies from background by fractal methods. *J Geochem Explor* 51:109–130
- Cheng Q, Bonham-Carter GF, Raines GL (2001) GeoDAS: a new GIS system for spatial analysis of geochemical data sets for mineral exploration and environmental assessment. The 20th Intern. Geochem. Explor. Symposium (IGES). Santiago de Chile, 6/5–10/5, pp 42–43
- Coordination de l'information sur l'Environnement Land Cover (Corine Land Cover, Italy) (2012) <http://www.pcn.minambiente.it/geoportal/catalog/search/resource/details.page?uuid=%7BA016A835-D5D8-4089-A9AE-2CC2D74B0C81%7D>
- Corona P, Barbati A, Tomao A, Bertani R, Valentini R, Marchetti M, Perugini L (2012) Land use inventory as framework for environmental accounting: an application in Italy. *IForest-Biogeosciences and Forestry* 5(4):204–209. <https://doi.org/10.3832/for0625-005>
- Costantini EAC, Dazzi C (2013) The soils of Italy. Springer, Berlin, p 354
- Courtney KD (1979) Hexachlorobenzene (HCB): a review. *Environ Res* 20:225–266
- Dearth MA, Hites RA (1991) Complete analysis of technical chlordane using negative ionization mass spectrometry. *Environ Sci Technol* 25:245–254
- DEFRA (Department for Environment, Food and Rural Affairs) (2011) Guidelines for environmental risk assessment and management: green leaves III. Defra and the collaborative Centre of Excellence in understanding and managing natural and environmental risks. Cranfield University, Bedford, p 84
- DoE (Department of Environment) (1995) A guide to risk assessment and risk management for environmental protection. HMSO, London
- Donnarumma L, Pompei V, Faraci A, Conte E (2009) Dieldrin uptake by vegetable crops grown in contaminated soils. *J Environ Sci Health B* 44:449–454
- Endosulfan Preliminary Dossier (2003) Umweltbundesamt, Berlin, pp 55
- ESRI (Environmental Systems Research Institute) (2012) ArcGIS desktop: release 10. Environmental Systems Research Institute, Redlands
- Estellano VH, Pozo K, Harner T, Corsolini S, Focardi S (2012) Using PUF disk passive samplers to simultaneously measure air concentrations of persistent organic pollutants (POPs) across the Tuscany Region, Italy. *Atmos Pollut Res* 3:88–94
- Eurostat (2014) http://ec.europa.eu/eurostat/statistics-explained/index.php/Pesticide_sales_statistics.explained/index.php/Pesticide_sales_statistics#Data_sources_and_availability
- Falandysz J, Brudnowska B, Kawano M, Wakimoto T (2001) Polychlorinated biphenyls and organochlorine pesticides in soils from the southern part of Poland. *Arch Environ Contam Toxicol* 40:173–178
- Fan MA, Alexeeff GV (1999) Public health goal for endrin in drinking water. Office of Environmental Health and Hazard Assessment. Environmental Protection Agency, California, pp 5–6
- Fang Y, Nie Z, Die Q, Tian Y, Liu F, He J, Huang Q (2017) Organochlorine pesticides in soil, air, and vegetation at and around a contaminated site in southwestern China: concentration, transmission, and risk evaluation. *Chemosphere* 178:340–349. <https://doi.org/10.1016/j.chemosphere.2017.02.151>
- Filzmoser P, Hron K (2008) Outlier detection for compositional data using robust methods. *Math Geosci* 40(3):233–248
- Filzmoser P, Hron K, Reimann C (2009) Principal component analysis for compositional data with outliers. *Environmetrics* 20(6):621–632
- Filzmoser P, Hron K, Tolosana-Delgado R (2012) Interpretation of multivariate outliers for compositional data. *Comput Geosci* 39:77–85
- Gabriel KR (1971) The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58(3):453–467
- Hron K, Templ M, Filzmoser P (2010) Imputation of missing values for compositional data using classical and robust methods. *Comput Stat Data Anal* 54(12):3095–3107
- ISPRA (2014a) Audizione dell'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) presso la Commissione Agricoltura, congiuntamente con la Commissione Ambiente, della Camera sul consumo di suolo, Audizione, Roma, 27/2/2014. [Italian]
- ISPRA (2014b) Il Consumo di suolo in Italia. Retrieved from <http://www.isprambiente.gov.it/it/publicazioni/rapporti/il-consumo-di-suolo-in-italia> [in Italian]
- ISTAT (2013) Indicatori ambientali urbani, Anno 2011. ISTAT, Roma [Italian]
- ISTAT (2016) Resident population on 1st January, 2017. <https://www.istat.it/en/population-and-households>
- Jayaraj R, Megha P, Sreedev P (2016) Organochlorine pesticides, their toxic effects on living organisms and their fate in the environment. *Interdiscip Toxicol* 9(3–4):90–100. <https://doi.org/10.1515/intox-2016-0012>
- Jennings A, Li Z (2014) Scope of the worldwide effort to regulate pesticide contamination in surface soils. *J Environ Manag* 146:420–443
- Jia H, Liu L, Sun Y, Sun B, Wang D, Su Y, Kannan K, Li Y-F (2010) Monitoring and modelling endosulfan in Chinese surface soil. *Environ Sci Technol* 44:9279–9284
- Jiang YF, Wang XT, Jia Y, Wang F, Wu MH, Sheng GY, Fu JM (2009) Occurrence, distribution and possible sources of organochlorine pesticides in agricultural soil of Shanghai, China. *J Hazard Mater* 170: 989–997
- Kim KH, Kabir E, Jahan SA (2017) Exposure to pesticides and the associated human health effects. *Sci Total Environ* 575:525–535

- Legislative Decree 152/2006 Decreto Legislativo 3 aprile (2006) n. 152, “Norme in materia ambientale”. Gazzetta Ufficiale n. 88 14-4-2006, Suppl Ord n. 96
- Li XH, Wang W, Wang J, Cao XL, Wang XF, Liu JC (2008) Contamination of soils with organochlorine pesticides in urban parks in Beijing, China. *Chemosphere* 70:1660–1668
- Lima, A., De Vivo, B., Cicchella, D., Cortini, M., Albanese, S., (2003). Multifractal IDW interpolation and fractal filtering method in environmental studies: an application on regional stream sediments of Campania Region (Italy). *Appl Geochem* 18 (12), 1853–1865. Doi: [https://doi.org/10.1016/S0883-2927\(03\)00083-0](https://doi.org/10.1016/S0883-2927(03)00083-0)
- Mackay D, Shiu WY, Ma KC (1992) Illustrated handbook of physical–chemical properties and environmental fate for organic compounds. Lewis Publishers, Chelsea
- Manz M, Wenzel KD, Dietze U (2001) Persistent organic pollutants in agricultural soils of central Germany. *Sci Total Environ* 277:187–198
- Maronna R, Martin R, Yohai V (2006) Robust statistics: theory and methods. John Wiley, p 436 ISBN: 978-0-470-01092-1
- Metcalfe RL (1995) Insect control technology. In: Kroschwitz J, Howe-Grant M (eds) Kirk-Othmer encyclopedia of chemical technology, vol 14. John Wiley and Sons, Inc., New York, pp 524–602
- Mishra K, Sharma RC, Kumar S (2012) Contamination levels and spatial distribution of organochlorine pesticides in soils from India. *Ecotoxicol Environ Saf* 76:215–225
- Moockel C, Macleod M, Hungerbühler K, Jones KC (2008) Measurement and modelling of diel variability of polybrominated diphenyl ethers and chlordanes in air. *Environ Sci Technol* 42:3219–3225
- Nizzetto L, Cassani C, Di Guardo A (2006) Deposition of PCBs in mountains: the forest filters effect of different forest ecosystem types. *Ecotoxicol Environ Saf* 63:75–83
- Oppolzer W (1991) Intermolecular Diels-Alder reactions. In: Trost BM, Fleming I (eds) Comprehensive organic synthesis, vol 5. Pergamon Press, Oxford chap 4.1
- Otero N, Tolosana-Delgado R, Solera A, Pawlowsky-Glahn V, Canals A (2005) Relative vs. absolute statistical analysis of compositions: a comparative study of surface waters of a Mediterranean river. *Water Res* 39:1404–1414
- Pacyna JM, Breivik K, Münch J, Fudala J (2003) European atmospheric emissions of selected persistent organic pollutants, 1970–1995. *Atmos Environ* 37:S119–S131
- PANNA (2008) Endosulfan monograph <http://www.panna.org/files/PAN%20Int%20Endosulfan%20Monograph.pdf>
- Parlar H, Hustert K, Gab S, Korte F (1979) Isolation, identification, and chromatographic characterization of some chlorinated C10 hydrocarbons in technical chlordane. *J Agric Food Chem* 27:278–283
- Pawlowsky-Glahn V, Buccianti A (2011) Compositional data analysis: theory and applications. John Wiley & Sons, p 400
- Pfaender FK, Alexander M (1972) Extensive microbial degradation of DDT in vitro and DDT metabolism by natural communities. *J Agric Food Chem* 20:842–846
- Pison G, Rousseeuw PJ, Filzmoser P, Croux C (2003) Robust factor analysis. *J Multivar Anal* 84:145–172
- Pozo K, Palmeri M, Palmeri V, Estellano VH, Mulder MD, Efstathiou CI, Sara GL, Romeo T, Lammel G, Focardi S (2016) Assessing persistent organic pollutants (POPs) in the Sicily Island atmosphere, Mediterranean, using PUF disk passive air samplers. *Environ Sci Pollut Res* 23(20):796–804
- Prapamontol T, Stevenson D (1991) Rapid method for the determination of organochlorine pesticides in milk. *J Chromatogr* 552:249–257
- Purnomo AS, Putra SR, Shimizu K, Kondo R (2014) Biodegradation of heptachlor and heptachlor epoxide-contaminated soils by white-rot fungal inocula. *Environ Sci Pollut Res* 21(19):11305–11312
- Qiu X, Zhu T, Yao B, Hu J, Hu S (2005) Contribution of dicofol to the current DDT pollution in China. *Environ Sci Technol* 39:4385–4390
- Qu C, Albanese S, Chen W, Lima A, Doherty AL, Piccolo A, Arienzo M, Qi S, De Vivo B (2016) The status of organochlorine pesticide contamination in the soils of the Campanian Plain, southern Italy, and correlations with soil properties and cancer risk. *Environ Pollut* 216:500–511
- Qu C, Albanese S, Lima A, Li J, Doherty AL, Qi S, De Vivo B (2017) Residues of hexachlorobenzene and chlorinated cyclodiene pesticides in the soils of the Campanian Plain, southern Italy. *Environ Pollut* 231(2):1497–1506. <https://doi.org/10.1016/j.envpol.2017.08.100>
- Reimann C, Filzmoser P, Garrett R (2002) Factor analysis applied to regional geochemical data: problems and possibilities. *Appl Geochem* 17(3):185–206
- Reimann C, Birke M, Demetriades A, Filzmoser P, O'Connor P, GEMAS Team (eds) (2014) Chemistry of Europe’s agricultural soils—part A: methodology and interpretation of the GEMAS data set. *Geologisches Jahrbuch (Reihe B)*, Schweizerbart, Hannover, p 528
- Senthil Kumar K, Kannan K, Subramanian AN, Tanabe S (2001) Accumulation of organochlorine pesticides and polychlorinated biphenyls in sediments, aquatic organisms, birds and bird eggs and bat collected from South India. *Environ Sci Pollut Res* 8:35–47
- Silva MH, Carr WC Jr (2009) Human health risk assessment of endosulfan. II: Dietary exposure assessment. *Regul Toxicol Pharmacol* 56: 18–27
- Stockholm Convention on Persistent Organic Pollutants (POPs) UNEP: persistent organic pollutants (2001) http://www.pops.int/documents/convtext/convtext_en.pdf
- Stockholm Convention on Persistent Organic Pollutants (POPs) (2005) World wide found, Stockholm convention “new POPs”: screening additional POPs candidates, April 2005, pp 38
- Stockholm Convention on Persistent Organic Pollutants (POPs) (2011) <http://chm.pops.int/>
- Suchan P, Pulkrabová J, Hajslová J, Kocourek V (2004) Pressurized liquid extraction in determination of polychlorinated biphenyls and organochlorine pesticides in fish samples. *Anal Chim Acta* 520: 193–200
- Szeto SY, Price PM (1991) Persistence of pesticide residues in mineral and organic soils in the Fraser valley of British Columbia. *J Agric Food Chem* 39:1679–1684
- Templ M, Hron K, Filzmoser P, (2011) Robcompositions: robust estimation for compositional data. Manual and Package, Version 1.4.4
- Thiombane M, Zuzolo D, Cicchella D, Albanese S, Lima A, Cavaliere M, De Vivo B (2018) Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model. *J Geochem Explor* 189:85–99
- UNECE (United Nation Economic Commission for Europe) 2010. Convention on long-range transboundary air pollution. http://www.unece.org/env/lrtap/pops_h1.htm
- US EPA (1991) Risk assessment guidance for superfund: volume I—human health evaluation manual (part B, development of risk-based preliminary remediation goals). U.S. Environmental Protection Agency, Washington, DC [EPA/540/R-92/003]
- US EPA (2000) Ultrasonic extraction, test methods for evaluating solid waste, method 3550C, revision 3. US Environmental Protection Agency, Washington, DC
- US EPA (2007) Endosulfan. Hazard characterization and endpoint selection reflecting receipt of a developmental neurotoxicity study and subchronic neurotoxicity study. PC Code: 079401. DP Barcode D338576. Reaves, E., Washington, DC: Office of Prevention, Pesticides and Toxic Substances, US Environmental Protection Agency. April 7, 2007. <http://www.epa.gov/pesticides/reregistration/endosulfan/>
- Van Den Boogaart KG, Tolosana-Delgado R, Bren R (2011) Compositions: analysis. R Package Version 1. pp. 10–12. Available at: <http://CRAN.R-project.org/package=compositions>
- Villanneau EJ, Saby NPA, Marchant BP, Jolivet C, Boulonne L, Caria G, Barriuso E, Bispo A, Briand O, Arrouays D (2011) Which persistent organic pollutants can we map in soil using a large spacing

- systematic soil monitoring design? A case study in Northern France. *Sci Total Environ* 409:3719–3731
- Walker K, Vallero DA, Lewis RG (1999) Factors influencing the distribution of lindane and other hexachlorocyclohexane in the environment. *Environ Sci Technol* 33:4373–4378
- Weinberg J (1998) Overview of POPs and need for a POPs treaty. Public forum on persistent organic pollutants—the international POPs elimination network
- WHO (2003) Health risks of persistent organic pollutants from long-range transboundary air pollution. http://www.euro.who.int/data/assets/pdf_file/0009/78660/e78963.pdf. pp 252
- Yadav IC, Devi NL, Li J, Zhang G, Breivik K (2017) Possible emissions of POPs in plain and hilly areas of Nepal: implications for source apportionment and health risk assessment. *Environ Pollut* 220: 1289–1300. <https://doi.org/10.1016/j.envpol.2016.10.102>
- Yang XL, Wang SS, Bian YR, Chen F, Yu GF, Gu CG (2008) Dicofol application resulted in high DDTs residue in cotton fields from Northern Jiangsu province, China. *J Hazard Mater* 150:92–98
- Yang D, Qi SH, Zhang JQ, Tan LZ, Zhang JP, Zhang Y, Xu F, Xing XL, Hu Y, Chen W, Yang JH, Xu MH (2012) Residues of organochlorine pesticides (OCPs) in agricultural soils of Zhangzhou City, China. *Pedosphere* 22:178–189
- Zhang ZL, Huang J, Yu G et al (2004) Occurrence of PAHs, PCBs and organochlorine pesticides in Tonghui River of Beijing, China. *Environ Pollut* 130:249–261
- Zhou Q, Wang J, Meng B, Cheng J, Lin G, Chen J, Zheng D, Yu Y (2013) Distribution and sources of organochlorine pesticides in agricultural soils from central China. *Ecotoxicol Environ Saf* 93:163–170